2017

Identifying Otosclerosis with Aural Acoustical Tests of Absorbance, Group Delay, Acoustic Reflex Threshold, and Otoacoustic Emissions

Douglas H. Keefe
Boys Town National Research Hospital, Douglas.Keefe@boystown.org

Kelly L. Archer
University of Nebraska - Lincoln

Kendra Schmid
University of Nebraska Medical Center, kkschmid@unmc.edu

Dennis F. Fitzpatrick
Boys Town National Research Hospital

M. Patrick Feeney
Veterans Administration and Oregon Health & Science University

See next page for additional authors

Authors
Douglas H. Keefe, Kelly L. Archer, Kendra Schmid, Dennis F. Fitzpatrick, M. Patrick Feeney, and Lisa L. Hunter

J Am Acad Audiol 28:838–860 (2017)

Identifying Otosclerosis with Aural Acoustical Tests of Absorbance, Group Delay, Acoustic Reflex Threshold, and Otoacoustic Emissions

DOI: 10.3766/jaaa.16172

Douglas H. Keefe*
Kelly L. Archer*†
Kendra K. Schmid‡
Denis F. Fitzpatrick*
M. Patrick Feeney§
Lisa L. Hunter**

Abstract

Background: Otosclerosis is a progressive middle-ear disease that affects conductive transmission through the middle ear. Ear-canal acoustic tests may be useful in the diagnosis of conductive disorders. This study addressed the degree to which results from a battery of ear-canal tests, which include wideband reflectance, acoustic stapedius muscle reflex threshold (ASRT), and transient evoked otoacoustic emissions (TEOAEs), were effective in quantifying a risk of otosclerosis and in evaluating middle-ear function in ears after surgical intervention for otosclerosis.

Purpose: To evaluate the ability of the test battery to classify ears as normal or otosclerotic, measure the accuracy of reflectance in classifying ears as normal or otosclerotic, and evaluate the similarity of responses in normal ears compared with ears after surgical intervention for otosclerosis.

Research Design: A quasi-experimental cross-sectional study incorporating case control was used. Three groups were studied: one diagnosed with otosclerosis before corrective surgery, a group that received corrective surgery for otosclerosis, and a control group.

Study Sample: The test groups included 23 ears (13 right and 10 left) with normal hearing from 16 participants (4 male and 12 female), 12 ears (7 right and 5 left) diagnosed with otosclerosis from 9 participants (3 male and 6 female), and 13 ears (4 right and 9 left) after surgical intervention from 10 participants (2 male and 8 female).

Data Collection and Analysis: Participants received audiometric evaluations and clinical immittance testing. Experimental tests performed included ASRT tests with wideband reference signal (0.25–8 kHz), reflectance tests (0.25–8 kHz), which were parameterized by absorbance and group delay at ambient pressure and at swept tympanometric pressures, and TEOAE tests using chirp stimuli (1–8 kHz). ASRTs were measured in ipsilateral and contralateral conditions using tonal and broadband noise activators. Experimental ASRT tests were based on the difference in wideband-absorbed sound power before and after presenting the activator. Diagnostic accuracy to classify ears as otosclerotic or normal was quantified by the area under the receiver operating characteristic curve (AUC) for univariate and multivariate reflectance tests. The multivariate predictor used a small number of input reflectance variables, each having a large AUC, in a principal components analysis to create independent variables and followed by a logistic regression procedure to classify the test ears.

Results: Relative to the results in normal ears, diagnosed otosclerosis ears more frequently showed absent TEOAEs and ASRTs, reduced ambient absorbance at 4 kHz, and a different pattern of tympanometric

*Boys Town National Research Hospital, Omaha, NE; †University of Nebraska Lincoln, Lincoln, NE; ‡Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE; §National Center for Rehabilitative Auditory Research, Veterans Administration and Oregon Health & Science University, Portland, OR; **Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Corresponding author: Douglas H. Keefe, Boys Town National Research Hospital, Omaha, NE 68131; Email: Douglas.Keefe@boystown.org

NIH grants R01 DC010202 and P30 DC004662.

A preliminary oral presentation of results was given at the May 2014 meeting of the Acoustical Society of America in Providence, RI.

absorbance and group delay (absorbance increased at 2.8 kHz at the positive-pressure tail and decreased at 0.7–1 kHz at the peak pressure, whereas group delay decreased at positive and negative-pressure tails from 0.35–0.7 kHz, and at 2.8–4 kHz at positive-pressure tail). Using a multivariate predictor with three reflectance variables, tympanometric reflectance (AUC = 0.95) was more accurate than ambient reflectance (AUC = 0.88) in classifying ears as normal or otosclerotic.

Conclusions: Reflectance provides a middle-ear test that is sensitive to classifying ears as otosclerotic or normal, which may be useful in clinical applications.

Key Words: absorbance, acoustic stapedius muscle reflex, ear canal, group delay, middle ear, otosclerosis, reflectance, transient evoked otoacoustic emission, tympanometry

Abbreviations: AC = air-conduction; ASR = acoustic stapedius muscle reflex; ASRT = acoustic stapedius muscle reflex threshold; AUC = area under the ROC curve; BBN = broadband noise; BC = bone-conduction; LMM = linear mixed model; NR = no response; OAE = otoacoustic emission; PC = principal component; PCA = principal component analysis; peSPL = peak-to-peak equivalent sound pressure level; Pnt = negative-tail pressure; Ppt = positive-tail pressure; ROC curve = receiver operating characteristic curve; SD = standard deviation; SE = standard error of the mean; SNR = signal to noise ratio; SPL = sound pressure level; TEOAE = transient evoked otoacoustic emission; TM = tympanic membrane; TPP = tympanometric peak pressure; WAI = wideband acoustic immittance; WB = wideband

INTRODUCTION

The etiology of middle-ear dysfunction including otosclerosis is often unknown without further and more invasive investigation. In patients with otosclerosis, signs of the disorder may appear after a thorough case history and an assessment of features observed on the audiogram. However, the audiogram cannot stand alone in differential diagnosis of otosclerosis (Rappaport and Provencal, 2002). Otosclerosis may be accurately diagnosed through a computed tomography scan in conjunction with exploratory surgery. Recently, aural acoustic tests including reflectance have provided evidence of their potential value in diagnosing middle ear pathologies including otosclerosis. This study evaluates the use of reflectance tests in identifying otosclerosis, in conjunction with using related aural acoustic tests including acoustic stapedius muscle reflex threshold (ASRT) and transient evoked otoacoustic emission (TEOAE) tests to classify ears as having an elevated risk for otosclerosis.

Background

Otosclerosis is a progressive and primary metabolic bone disease that results from abnormal otic capsule bone remodeling and typically manifests in early adulthood. Initially, otosclerosis presents with resorption of bone and replacement with soft, highly vascularized bone. The second stage is characterized by bony fixation of the anterior stapes footplate and then progresses to ankylosis of the entire footplate, producing a conductive or mixed hearing loss in the range up to 60 dB (Merchant and Nadol, 2010). The resulting hearing loss can progress and a sensorineural component may be observed in later stages (Cureoglu et al, 2010). The disease is genetically mediated via autosomal dominant transmission. Many clinical cases have been reported to be arbitrary, i.e., without a known genetic cause. Clinical otosclerosis is a major origin of acquired hearing loss in the adult population. Thys and Van Camp (2009) report that otosclerosis develops in 1% of the population and is found more often in women than men, whereas Hannula et al (2012) report that it occurs in 1.3% of adults. Thus, it is imperative that this disorder be identified accurately and in a timely manner so that appropriate intervention can be provided.

Aural Acoustic Tests to Screen for and Diagnose Otosclerosis

A focus of otosclerosis research concerns the inability to consistently differentiate this disease from normal ears or other pathologies affecting the middle ear. Cases of otosclerosis are suspected based on frequent patient reports of tinnitus, hearing concerns, a positive family history for otosclerosis, and specific audiometric features. These features include air-bone gaps, presence of a Carhart notch, normal or reduced tympanometric mobility, and absent acoustic reflexes. However, these traits can be observed individually or may not be evident at all in the early stages. For example, an air-bone gap is frequently seen in otosclerosis, but also in other pathologies. The Carhart notch is not a definitive indication of otosclerosis as other middle-ear pathologies alter the normal resonance characteristics of the ossicular chain and result in the smallest air-bone gaps near 2000 Hz (Shanks and Shohet, 2009).

Otosclerosis is not the only disorder to produce the previously mentioned characteristic features on the audiogram. After identification of suspected otosclerosis, patients are referred for an additional evaluation, which may include tuning fork tests, computerized tomographic scanning of the temporal bone, and eventual

exploratory surgery. The nonsurgical procedures are time consuming and expensive, and may not lead to an accurate diagnosis. Furthermore, it is not practical to employ these procedures on every patient that presents with a conductive hearing loss. At this time, no single diagnostic test can noninvasively and accurately identify otosclerosis. Studies are next described reporting acoustic test data in otosclerotic ears of adults.

Clinical Tympanometry

Clinical tympanometry, which is typically performed in adult ears based on the compensated admittance at 226 Hz, is ineffective in identifying pathologies such as otosclerosis, which do not directly affect the tympanic membrane (TM). As a group, otosclerotic ears show a reduced mean compensated admittance magnitude at 226 Hz compared with normal ears, but individually, there is substantial overlap with the normal-hearing ears (Shahnaz and Polka, 1997). Immittance assessments using multifrequency tympanometry are more accurate than 226-Hz tympanometry in identifying individuals with otosclerosis compared with ears with normal function (Shahnaz, Bork, et al, 2009).

Sweep Frequency Impedance

A sweep frequency impedance test is an aural acoustic test of middle-ear function that is concluded to be more sensitive to middle-ear stiffness changes associated with an adult otosclerotic ear than 226-Hz tympanometry (Zhao et al, 2002). It appears that the overall diagnostic accuracy of this test has not yet been assessed for classifying ears as normal or otosclerotic.

Absorbance and Group Delay

An aural acoustic test to measure pressure reflectance at the probe tip over a wide bandwidth of frequencies (typically 0.25–8 kHz or wider) was performed at ambient pressure in the ear canal to study middle-ear function in ears with normal hearing (Keefe et al, 1993; Voss and Allen, 1994). A corresponding tympanometric test of pressure reflectance was developed over the same wideband (WB) frequency range and over a range of tympanometric air pressures in the ear canal (Keefe and Levi, 1996; Margolis et al, 1999).

All information in the pressure reflectance of the ear is parameterized at each frequency in terms of the absorbance, which is one minus the magnitude squared of the pressure reflectance, and the group delay, which is the product of $(1/2\pi)$ times the gradient of the pressure reflectance phase with respect to frequency (Keefe et al, 1993; Voss and Allen, 1994). Absorbance varies between zero if none of the power of the sound stimulus is absorbed by the middle ear and one if all the power is absorbed. The group delay is approximately equal to the sum of the time for sound to travel from the probe to the TM and back plus the group delay of the middle ear measured at the TM. A tympanometric peak pressure (TPP) is defined in terms of the absorbance tympanogram averaged over low frequencies from 0.376 to 2 kHz. This allows a reduction in the number of variables describing the reflectance tympanogram by calculating the absorbance and group delay as functions of frequency at the TPP, as well as at the most positive and most negative pressures used in the tympanometric sweep. Normative adult absorbance and group-delay data along with test–retest values for both the ambient and tympanometric types of these reflectance tests are available (Feeney et al, 2016).

Aural acoustic reflectance tests show promise in studies to identify ears with otosclerosis relative to ears with normal function or ears with other types of middle-ear disorders. Some studies described in the later sections expressed results in terms of energy (or power) reflectance, and all prior reflectance studies of otosclerosis have focused only on the test at ambient pressure in the ear canal. To unify the present description, all energy or power reflectance results in these studies are described herein in terms of their corresponding absorbance.

Feeney, Grant, et al (2003) showed that absorbance in ears with otosclerosis was reduced compared with normal-hearing ears at frequencies below 1 kHz, and more sensitive to the presence of otosclerosis than 226 Hz tympanometry. Allen et al (2005) reported similar results in a single bilateral case of otosclerosis, as did Shahnaz and Bork (2006) for four otosclerotic ears. In a study in 15 ears with confirmed otosclerosis before and after stapes surgery, the ambient absorbance was smaller in presurgery ears compared with normal ears between 0.7 and 1 kHz, and increased in that frequency range after surgery (Shahnaz, Longridge, et al, 2009). Absorbance in postsurgery ears was increased by a smaller relative amount between 2 and 4 kHz compared with presurgery ears. Compared with normal ears, postsurgery ears trended to larger mean absorbance at frequencies below 0.5 kHz followed by smaller mean absorbance at higher frequencies up to 1.1 kHz. Shahnaz, Longridge, et al (2009) concluded that ambient absorbance might be useful to monitor the status of the stapes reconstructive surgery and to evaluate different surgical protocols. Consistent with these studies, Nakajima et al (2012) reported a decreased mean absorbance between 0.4 and 1 kHz in 14 ears with a fixed stapes due to otosclerosis relative to normal ears.

The abilities of multifrequency tympanometry and ambient absorbance to distinguish otosclerotic from normal ears were measured using an analysis of variance approach and a receiver operating characteristic (ROC) curve approach, using the area under the ROC

curve (AUC) as the measure of diagnostic accuracy (Shahnaz, Bork, et al, 2009). Absorbance was lower in the 28 otosclerotic ears, with no difference between male and female ears. The mean absorbance was significantly lower in otosclerotic ears at frequencies between 0.4 and 1 kHz, in agreement with earlier studies. ROC curve analyses were performed to distinguish otosclerotic and normal ears at the six third-octaves (0.315–1 kHz) containing this frequency range, with the largest AUC of 0.86 at 0.5 kHz. The diagnostic accuracy of the multifrequency tympanometry measure was equivalent to that of absorbance at 0.5 kHz, and both were more accurate than 226-Hz tympanometry.

Acoustic Stapedius Muscle Reflex (ASR) Responses

ASRs in ears with normal function were first described by Metz (1946) using the first impedance bridge developed for human measurements. Today, they are typically measured in the clinic in terms of a shift in the magnitude of the compensated admittance magnitude at 226 Hz in a quiet condition relative to a condition in which a second activator signal is presented at varying sound pressure levels (SPLs) in the ipsilateral ear (i.e., in the same ear as the reference tone) or in the contralateral ear. The ASRT is the lowest SPL at which a significant shift in the admittance magnitude is detected. The ASR may be measured using either tonal activator signals (typically at frequencies of 0.5, 1, or 2 kHz) or a broadband noise (BBN) activator.

By comparing acoustic impedance measurements at the probe tip in adult ears with normal hearing and otosclerosis, an ASR in normal ears shifted the mean values of the impedance to values characteristic of the mean impedance in otosclerotic ears (Feldman and Zwislocki, 1965). The ASR is absent in otosclerotic ears pre- and postsurgery, inasmuch as the stapedius tendon is typically, although not always, removed from the stapes just before the surgical repair.

A WB ASR test substitutes a WB reference stimulus in place of the 226 Hz tone to assess the ASR shifts across frequency (0.25–8 kHz) (Feeney and Keefe, 1999). A corresponding ASRT has been measured based on WB shifts in admittance, reflectance, and the sound power absorbed by the ear (Feeney and Keefe, 2001). The potential advantages of a WB technique over a single-frequency technique is to examine ASR shifts over the frequency range important for speech perception and to facilitate higher-frequency ASRT measurements in young infant ears in which ASR shifts are weak or absent at 226 Hz (Weatherby and Bennett, 1980). ASR responses were measured in the present study in terms of WB reflex shifts in absorbed sound power across frequency using procedures described in Keefe et al (2017). Hunter et al (2017) reported that ASRTs in newborn infant ears passing a newborn hearing screening examination were significantly elevated in an ambient ASRT test compared with a pressurized ASRT test at TPP. They found that pressurized WB ASRTs measured in the real ear to broad band noise were significantly elevated in newborn infant ears with conductive and sensorineural hearing loss, thus demonstrating the clinical value of these tests. In adult ears with normal hearing, WB ASRTs using BBN activators are approximately 12 dB more sensitive than the standard clinical method (Feeney, Keefe, et al, 2003). A corresponding comparison of WB and single-frequency ASR tests using tonal activators in the ipsilateral ear has been reported (Schairer et al, 2007), but no such comparison of contralateral ASRTs has been reported.

Wideband Acoustic Immittance (WAI) Responses

The term aural acoustic immittance refers to one or both of aural acoustic impedance and acoustic admittance transfer functions measured in the ear canal (ANSI, 2012), whether at ambient pressure in the ear canal, as part of a tympanometry test in which the ear-canal pressure is varied over one or more values, or in an ASR test measuring a change in acoustic impedance or admittance induced by the acoustic reflex. The term WAI is a generalization to describe measurements across a wide frequency range of aural acoustic transfer functions including impedance, admittance, or reflectance (Keefe, 2007). The absorbance, group delay, and ASR responses used in the present study form such a WAI test battery. This usage is consistent with a consensus statement on research needs and clinical applications of WAI (Feeney et al, 2013).

TEOAE Responses

TEOAEs, which were first discovered by Kemp (1978), are a by-product of the compressive nonlinearity of the function of outer hair cells in a normal cochlea. They are sensitive to cochlear function, but also sensitive to middle-ear function, even though they are generated within the cochlea. This is because the detection of a normal TEOAE in the ear canal depends on normal forward transmission of the sound stimulus energy from the ear canal through the middle ear to the cochlea and normal reverse transmission of the cochlear-generated otoacoustic emission (OAE) back through the middle ear to the ear canal. Using chirp stimuli, TEOAEs were measured in the present study as a nonlinear residual to frequencies up to 8 kHz (Keefe et al, 2016), which is higher than the maximum test frequency of 4–5 kHz in clinical TEOAE tests.

Otosclerosis patients typically have absent clinical ASR responses and absent or elevated evoked OAEs (Herzog et al, 2001; Zhao et al, 2003). It is also the case that some patients with normal middle-ear function have absent acoustic reflexes (Mukerji et al, 2010), or

have absent or elevated TEOAEs, thus these tests may result in false-positive results. An absent ASR or TEOAE in such cases means that no ASR or TEOAE response was detected by the given test.

Research Questions Addressed

A combination of aural acoustic tests may be more effective than any single test in identifying otosclerotic ears relative to ears with normal function. Although absent or weak ASR and TEOAE responses are consistent with otosclerosis, they may also be absent or weak in other types of conductive or cochlear pathologies. Specifically, TEOAEs and distortion product OAEs may be affected by Eustachian tube dysfunction resulting in positive or negative pressure (Trine et al, 1993; Plinkert et al, 1994; Sun and Shaver, 2009). Thus, a test of middle-ear function other than an evoked OAE or ASR test is needed to better identify a risk of otosclerosis. The ambient reflectance test has the potential for absorbance to differentiate between conductive and sensorineural dysfunction, and differentiate between differing types of middle-ear pathology in human temporal bones (Feeney, Grant, et al, 2003; Merchant et al, 2016) and in patients with conductive impairments (Nakajima et al, 2012), but no data are available on pressurized forms of these tests in ears with otosclerosis. The present research examined the relative abilities of absorbance and group delay to classify ears as having otosclerosis or normal hearing. These tests were performed at ambient pressure and at varying tympanometric pressure in the ear canal, which enabled a comparison between ambient and tympanometric reflectance tests.

Mean differences between responses in normal and otosclerotic ears were assessed as a linear mixed model (LMM) described in "Methods." The analysis groups included normal ears (Normal), ears diagnosed with otosclerosis (Otosclerosis), and ears have receiving surgical intervention for otosclerosis (Surgery). As a sidearm study, the ASRTs were compared in normal ears under ipsilateral and contralateral conditions using both tonal and BBN activators in the single-frequency (Clinical) and WB (Research) tests. It is hypothesized that ASRTs are similar for Research and Clinical acoustic reflex tests, for Otosclerosis and Surgery groups, and for ipsilateral and contralateral testing modes. It is further hypothesized that the mean absorbance and group delay responses differ between ambient and TPP conditions, and that mean absorbance and group delay are the same in Surgery and Otosclerosis groups, with reduced absorbance and decreased group delay compared with the Normal group. The hypothesized reductions in group delay and absorbance at low frequencies in otosclerotic ears would result from increased stiffness at the TM due to increased stiffness of the annular ligament in otosclerotic ears relative to normal ears (Zwislocki, 1962). Diagnostic accuracies of various tests were measured using ROC analyses of both univariate and multivariate classifiers, in which the input variables were combined in the latter classifier across frequency and response type. Reflectance results were interpreted using a test battery that also included TEOAEs and ASRTs to confirm the functional status of each test ear.

METHODS

The research plan for using human participants was approved by the Institutional Review Boards at Boys Town National Research Hospital (Omaha, NE) and the University of Nebraska–Lincoln. After the completion of paperwork, each participant completed several tests during a single, two-hour session in a laboratory room at Boys Town National Research Hospital.

Clinical Procedures

All clinical tests were performed within a sound-attenuated booth using ER3A insert earphones. All adult participants received air-conduction (AC) and bone-conduction (BC) audiometry (GSI 61 audiometer). The AC test frequencies included 0.25, 0.5, 1, 1.5, 2, 3, 4, 6, and 8 kHz. The BC test frequencies included 0.25, 0.5, 1, 2, and 4 kHz, so that air-bone gap data were available at all BC test frequencies. Participants also received a clinical 226-Hz admittance tympanometry test and clinical ASR tests (GSI Tympstar).

Clinical audiometric and immittance assessments of each participant included (a) AC audiometry, (b) BC audiometry, (c) 226-Hz tympanometry, and (d) clinical ASR measurements in ipsilateral and contralateral test modes, using tonal reflex activators of 0.5, 1, and 2 kHz in both modes, and a BBN activator in ipsilateral mode. For each activator type, the clinical ASRT (dB SPL) was determined as the lowest SPL that produced at least a 0.03 mmho change in admittance magnitude (or a 0.02 mmho change in a few ear tests, with 1 mmho equal to 0.001 mho in centimeter–gram–second units) on at least one ascending run. ASRTs in this study were limited to levels of 105 dB SPL for tonal activators and 80 dB SPL for the BBN activator for safety reasons (Hunter et al, 1999). An absent reflex response at all activator levels was coded as no response, and scored by a numerical value of 110 dB SPL for tonal activators and 85 dB SPL for BBN. This numerical coding of absent response facilitated interval comparisons of ASRTs in all test ears.

Participants

Twenty-five adults between the ages of 19 and 63 yr participated in this study. Each test ear was categorized

in one of three participant groups: Normal, Otosclerosis, and Surgery. It was possible for a participant to have either one or both ears in a group. Table 1 shows the mean age and gender distributions across groups.

Normal Group

The Normal hearing group included 23 ears in 16 participants. The inclusion criteria for the Normal hearing group were as follows: (a) no significant history of middle-ear pathology or otologic surgery, (b) otoscopically clear external auditory canals, (c) pure-tone AC and BC thresholds no >15 dB HL from 0.25 to 8 kHz, (d) air-bone gaps no >10 dB between frequencies of 0.25–4 kHz, and (e) normal clinical tympanometry at 226 Hz, defined as peak compensated static acoustic admittance between 0.3–1.5 mmho, peak pressures within ±50 daPa, and ipsilateral and contralateral acoustic reflexes essentially intact between the levels of 80 and 105 dB SPL for tonal activators and up to 80 dB SPL for the BBN activator.

Otosclerosis Group

The diagnostic Otosclerosis group included nine participants, 12 ears, who did not pursue surgical intervention. The inclusion criteria for the Otosclerosis group were: (a) otoscopically clear external auditory canals, and (b) either diagnosis of otosclerosis by an ear, nose, and throat physician based on findings of exploratory surgery, or suspected diagnosis of otosclerosis by the physician based on audiometric data and otologic examination with no prior history of surgery for otosclerosis.

Surgery Group

Participants in the otosclerosis Surgery group were recruited based on history of prior surgical intervention for otosclerosis. The Surgery group included 10 participants and 13 ears. Eleven ears received a stapedotomy, which is a surgical procedure that required a small prosthesis be inserted into a small hole that is created, via laser or manually, in the central footplate of the stapes. One ear received a stapedectomy, which involved complete or partial removal of the stapes footplate with a micro prosthesis to replace the diseased stapes bone (Fisch, 2009). A final ear received surgery, but the type of intervention was unknown. Surgery to treat otosclerosis does not typically preserve the stapedius tendon, so that each test ear in the Surgery group was expected to have an absent ASR.

Table 1. Age Distribution and Gender Profile of Human Participant Groups (SD)

Participant Characteristics	Normal	Otosclerosis	Surgery
N (ears)	23	12	13
N (left/right ears)	10/13	5/7	9/4
N (female, male, all participants)	12, 4, 16	6, 3, 9	8, 2, 10
Mean age (yr)	29	49	44
SD (yr)	9	10	12
Median age (yr)	28	51	48
Maximum age (yr)	51	63	65
Minimum age (yr)	19	28	27

Instrumentation and Research Tests

The research test battery included ambient and tympanometric reflectance, WB ASRT and TEOAE tests. All testing was conducted with the participant comfortably seated in a sound-attenuated booth. Research data were acquired using custom software on a Windows computer with two-channel sound card (CardDeluxe) with 24-bit resolution and bidirectional serial port (RS-232). All measurements used the same ear probe (Titan, Interacoustics), which had two receiver ports to deliver sound stimuli and one microphone port to measure acoustic pressure. One receiver generated sound stimuli under the control of a digital-to-analog converter channel one on the sound card, and the microphone output was sent to an analog-to-digital converter on the sound card. The sample rate per channel was 22.05 kHz. An additional port exiting the probe tip transmitted air-pressure changes generated by the pump in the tympanometer (AT235, Interacoustics, with modified firmware). A pressure sensor within the tympanometer cable measured the actual pressure, and a pressure controller circuit matched the measured air pressure to the desired pressure. During any pressurized measurement, air pressure was varied in the ear canal, and the tympanometer communicated the measured air pressure back to the computer using a bidirectional serial port.

Reflectance Test

The system was calibrated for reflectance measurements (Liu et al, 2008), which used a click stimulus (0.2–8 kHz) output by digital-to-analog converter channel one, by recording the pressure responses in a pair of hard-walled cylindrical tubes (with lengths of 292 and 8.2 cm). Each tube was closed at its opposite end with cross-sectional internal diameter of 7.94 mm, similar to that of an average adult ear. The incident pressure spectrum and the source reflectance of the system probe were calculated using these measured data and an acoustical model of sound propagation in cylindrical tubes with viscothermal wall loss. This calibration was performed daily before any ear test.

For the subsequent ear tests, the probe was placed into the ear canal with a leak-free insertion using a soft

plastic impedance probe tip. A repetitive sequence of clicks was generated by the probe to which the pressure response in the ear canal was measured using the microphone. All ear reflectance measurements were calculated in terms of these measured data along with the incident pressure and source reflectance stored from the most recent calibration for that day. The system also measured the acoustic admittance at the probe tip. The peak-to-peak equivalent sound pressure level (peSPL) of the click in the average adult ear was 92 dB. Preliminary measures were used to validate that no distortion was present at the click levels used for these reflectance (absorbance and group delay) measurements.

The ear-canal data in tests at ambient pressure in the ear canal consisted of 32 responses to clicks collected over 2 sec. The cross-sectional area of the ear canal at the probe tip was acoustically estimated in terms of the admittance at the probe tip. The pressure reflectance of the ear was calculated as a function of frequency in terms of this admittance and the estimated area, and this reflectance was represented in terms of the absorbance and group delay. The length of the ear canal along its central midline between probe tip and TM was acoustically estimated using the absorbance and group delay at frequencies above 2 kHz (Keefe et al, 2015).

The tympanometric reflectance test was similar to the ambient reflectance test in that responses were recorded to a sequence of WB clicks. The additional variable was that the tympanometric pressure was swept over the desired range during the presentation of the clicks, so that each click response was recorded at a slightly different ear-canal pressure. Two reflectance tympanograms were measured using an ear-canal pressure that was either downswept starting at +200 daPa down to −300 daPa, or upswept starting from −300 daPa up to +200 daPa. The TPP was calculated for each sweep direction by calculating the pressure at which the maximum occurred in a low-frequency absorbance averaged over frequencies from 0.376 to 2.0 kHz. The TPP of the ear was defined as the average of the TPPs from the downswept and upswept tympanograms, and this TPP was used in subsequent ASR testing. This average reduced any bias effects associated with the polarity of the tympanometric sweep.

The test battery began with the downswept reflectance tympanometry test, the ambient reflectance test, and the upswept reflectance tympanometry test. The downswept tympanometry test was completed first because it provided a strong test that the probe was inserted into the ear canal without any pressure leaks. This is based on the property that the ear-canal pressure was initially ramped up from ambient to 220 daPa (i.e., slightly above the maximum air pressure of 200 daPa at which data were analyzed), and it would be impossible to pressurize the ear canal to this maximum air pressure if there were any leak between the probe and the ear-canal wall at the entrance. Reflectance tympanometry data were represented as absorbance tympanograms and group delay tympanograms.

For individual ears, procedures to measure the ambient and tympanometric reflectances are fully described with examples (Keefe et al, 2015), which also describes the techniques to acoustically estimate the ear-canal area and length. Normative data on absorbance, group delay, ear-canal area, and length are available (Feeney et al, 2016).

ASR Test

The WB ASR test was performed in each test ear with the same probe insertion after all reflectance testing was completed, such that the ASR test data were acquired by first adjusting the air pressure in the ear canal to the TPP determined from the lowpass-filtered absorbance tympanograms. An ASR test differed according to the choice of activator signals from BBN or tonal activators at 0.5, 1, and 2 kHz and was performed in both ipsilateral and contralateral test modes.

The ASR pulsed-activator stimulus set included five clicks delivered by one receiver port, and four pulsed activators were delivered in an interleaved manner with the clicks by a second receiver port. Each click had a stimulus bandwidth from 0.2 to 8 kHz and was the same click used in the reflectance/admittance test. Each pulsed activator was of 116 msec duration with onset and offset ramps, and was presented between a pair of clicks. The duration of this pulsed stimulus set was 0.79 sec, with an additional 0.79 sec of silence between subsequent presentations to allow any ASR shift to be substantially attenuated. This resulted in an ipsilateral reflex test as both the reference signals (i.e., the clicks) and the signals activating the ASR (e.g., brief 0.5 kHz pulses) were delivered to the same ear into which the probe was inserted. Contralateral ASR testing was performed by inserting an ER3A earphone into the ear opposite to that in which the probe was inserted and by presenting the five clicks through the probe and the four pulsed activators through the contralateral earphone.

The initial or "baseline," click in each stimulus set measured the sound pressure response at the probe in the assumed absence of any ASR effect. The difference of each postactivator click with respect to this baseline click provided a set of four click-difference waveforms. A sufficiently large shift in any of these click-difference waveforms was associated with the presence of a measurable ASR.

All data were highpass filtered below 0.2 kHz to attenuate low-frequency noise. The stimulus set was repeated once at each of 10 activator levels in 5 dB increments of activators ascending in level. Tonal activator levels ranged from 60 to 105 dB SPL, whereas BBN

activator levels ranged from 35 to 80 dB SPL. Each activator level was specified by the SPL measured in a reference 2-cm^3 coupler, although the real-ear activator SPL was also measured for ipsilateral ASR tests in each ear. Air pressure in the ear canal was continuously monitored during the ASR test at TPP using a pressure sensor within the probe assembly. Whenever the measured air pressure deviated by >10 daPa from the desired TPP, the system paused the reflex test, adjusted the pump so as to produce the desired pressure, and then restarted the reflex test at the current activator level. The ASR was detected in terms of the click differences in the sound power absorbed by the ear in response to the baseline click and to each of the four subsequent clicks.

The ASRT was calculated based on the classification of the ASR response as present or absent at each activator level and the underlying similarity and consistency of the ASR shift data across the range of activator levels. When the ASR responses at all test activator levels were absent, the ASRT was coded as no response (NR) and assigned a numerical value 5 dB above the maximum activator SPL. That is, an NR was coded as 110 dB SPL (re: 2-cm^3 coupler) for tonal activators and 85 dB SPL for BBN activators. This placed an NR value on an interval scale of ASRT. Detailed methods used to measure the WB ASRT are fully described with examples in individual ears (Keefe et al, 2017), and normative WB ASRT data in adult ears are available (Feeney et al, 2016).

TEOAE Test

TEOAE testing at ambient pressure in the ear canal was performed in each ear after the reflectance and ASR tests using the same probe insertion. A so-called positive chirp stimulus was used to evoke the TEOAE response. The spectrum of the positive chirp had an approximately constant incident pressure magnitude (Goodman et al, 2009) at all frequencies between 0.5 and 8 kHz, and the chirp waveform swept at a rate of 174.6 Hz/msec from low to high frequencies over about 43 msec. The rationale for measuring TEOAEs with a chirp stimulus rather than a click stimulus is that, for the same total energy in the stimuli, the peSPL of the chirp is much lower than that of the click, for example, the peSPL of the chirp was 11.8 dB lower than that of the click in Keefe et al (2016). The chirp stimuli used in this study were identical to those used in Putterman et al (2017), for which the mean peSPL in normal adult ears was 76.4 dB. Because it is the peak levels that generate system distortion, the chirp stimulus is advantageous as it spreads out the energy in time, whereas a click stimulus concentrates the same energy in a brief time interval.

A nonlinear TEOAE residual was calculated using a double-evoked method (Keefe, 1998; Keefe and Ling, 1998) based on responses recorded to the positive chirp stimuli over a one-minute test duration. This residual is generated within the cochlea due to the compressive nonlinearity of cochlear mechanics, and the double-evoked method provides TEOAE spectral responses to higher frequencies than in the clinical test. The signal to noise ratio (SNR; in dB) of the TEOAE spectrum was analyzed over the bandwidth from 0.7 to 8 kHz and octave-averaged at center frequencies from 1 to 8 kHz. The TEOAE test procedures using this measurement system are described with normative adult data in Keefe et al (2016), and the performance of these test procedures in classifying ears with normal hearing or a sensorineural hearing loss are described in Putterman et al (2017).

TEOAEs were analyzed in the present study over frequency ranges centered at half-octave frequencies between 1 and 8 kHz in terms of the TEOAE signal level, noise level, and SNR, which was calculated as the larger of 0 dB or the difference in the TEOAE signal and noise levels.

Statistical Analyses

LMMs

An LMM approach allows for assessment of mean differences between groups, while considering both interaction and main effects, and accounting for the complex correlation structure present in this setting. For analyzing mean differences in ambient and (downswept) tympanometric reflectance and for analyzing mean differences in TEOAE signal level, noise level, and SNR, an LMM was used with one between-subjects factor group (Otosclerosis/Surgery/Normal) and one or two within-subject factors (type of measurement and frequency). A third within-subject factor, ear (left/right), was also used to structure the within-subject covariance matrix but was not used as a factor in any LMM because of the lack of any significant ear effect. Data were evaluated using a direct (Kronecker) product covariance structure based on the levels of the within-subject factors. The data on sex was not included as a factor because of an unbalanced distribution of test ears from women and men across the three participant groups. All other factors were treated as fixed effects, and their main effects and interactions were evaluated. Average measurements among the three levels of group were compared at each frequency and within type of measurement averaged over ear. To account for multiple comparisons between group means, a Hochberg adjustment was applied to the p values for the differences among means. All statistical significance tests were two-sided using a critical p value of 0.05 to determine statistical significance.

To further explain the magnitude of each effect, standardized effect sizes are presented. Each effect size in

the ASRT test, and the ambient and tympanometric absorbance and group delay tests, was calculated as the difference in mean values relative to their pooled standard deviation (SD). Effect sizes were characterized as either small up to 0.2, medium up to 0.5, or large above 0.8 (Cohen, 1992). Statistical analyses were generated with SAS/STAT software.

To analyze mean differences between clinical and WB ASRTs, seven types of ASRT tests were evaluated in a nested model within type contralateral (which had three levels to represent data at 0.5, 1, and 2 kHz tonal activators) and ipsilateral (which had the same three levels to represent data at 0.5, 1, and 2 kHz, and a fourth level for the BBN activator). No contralateral BBN test data were acquired in clinical ASR testing. One limitation was that there was no variation in ASRT data for some combinations of factors for the Surgery group level. This occurred whenever the SD of the ASRTs was zero, which corresponded to a ceiling effect of all Surgery ears having an NR on the ASRT test.

Although some violations to the normality assumption were present, ASRT data were analyzed using an LMM as it provided the best option to model the data considering the complex correlation structure in which test data were presented in one ear of some participants and both ears of others. With this lack of desirable variation in mind, an LMM remained an analysis option that could reveal differences in means, recognizing that the resulting p values have some inherent bias. The between-subject factor considered was group, and the two within-subject factors considered were ear (left/right) and ASRT-test-type (Clinical/Research). An additional within-subject factor was activator-ear (contralateral/ipsilateral), which had three activator factor levels for a contralateral ASRT test (activators using 0.5, 1.0, and 2.0 kHz tones) and four activator levels for an ipsilateral ASRT test (activators using 0.5, 1.0, and 2.0 kHz tones, and BBN). That is, Clinical and Research ASRTs were measured ipsilaterally and contralaterally in each participant group using tonal activators, but were only measured ipsilaterally in each participant group using the BBN activator. Owing to the complexity of the design, data were evaluated with an LMM that included a random effect for participants and correlations because of the within-subject factors. The complex covariance structure considered the correlated nature of the response within participants, for example, across both ears. The group, ASRT-test-type, activator-ear, and ear were entered as fixed effects. Main effects and interactions were evaluated. Differences between ears were not observed. However, because the interaction with group, ASRT-test-type, and activator-ear was significant in Results, average measurements were compared for ASRT-test-Type, paired values for contralateral and ipsilateral tests, and activator-ear.

AUC Analyses

In tests to classify ears as normal or impaired, the specificity is the proportion of normal ears (between 0 and 1) correctly classified as normal, and the sensitivity is the proportion of impaired ears (between 0 and 1) correctly classified as impaired. A test is perfectly accurate if all ears are correctly classified, for which the sensitivity and specificity are each equal to one. The ROC curve is a plot of the sensitivity versus one minus the specificity. The AUC is a summary statistic of the ROC curve that is equal to one for a perfect test and 0.5 for a test at random performance, in which each ear is classified with a probability of 0.5 to be either impaired or normal. In comparing multiple tests, the test with the largest AUC is the most accurate test.

The nonparametric AUC was calculated using the Mann–Whitney U-test statistic to assess the accuracy of reflectance data to classify ears as normal or otosclerotic. The AUC was calculated for each absorbance and group delay variable as a univariate classifier. These variables were selected from both ambient and tympanometric reflectance test data. A subsequent multivariate classifier was devised to classify ears as otosclerotic or normal, and is further described in the later sections. The ROC analyses were calculated using MATLAB (release 2016b) including its statistics and machine learning toolbox.

Comparison between LMM and AUC Analyses

The LMM assesses the magnitude of the difference between the test means of two or more groups, whereas the AUC is an unbiased statistic to quantify the accuracy with which a test correctly classifies a case into one of two classes (in this case Normal or Otosclerosis). These statistical tests provide complementary information on test performance. The LMM formally assumes that the test scores in each group are normally distributed, whereas the nonparametric AUC statistic does not require that assumption.

RESULTS

Audiometric Results

Figure 1 (top panel) shows the mean ± 1 SE of the mean (SE) AC audiograms in each group, and Figure 1 (bottom panel) shows the mean air-bone gaps. The mean HL and air-bone gap in the Normal group were within normal limits across frequency, as required by the inclusion criteria. The mean HL was most elevated in the Otosclerosis group in the range of 35–45 dB HL across frequency, and showed a maximum mean air-bone gap of 32 dB at 0.25 kHz, which remained larger than the 20 dB gap at 0.5 and 1 kHz. The mean air-bone

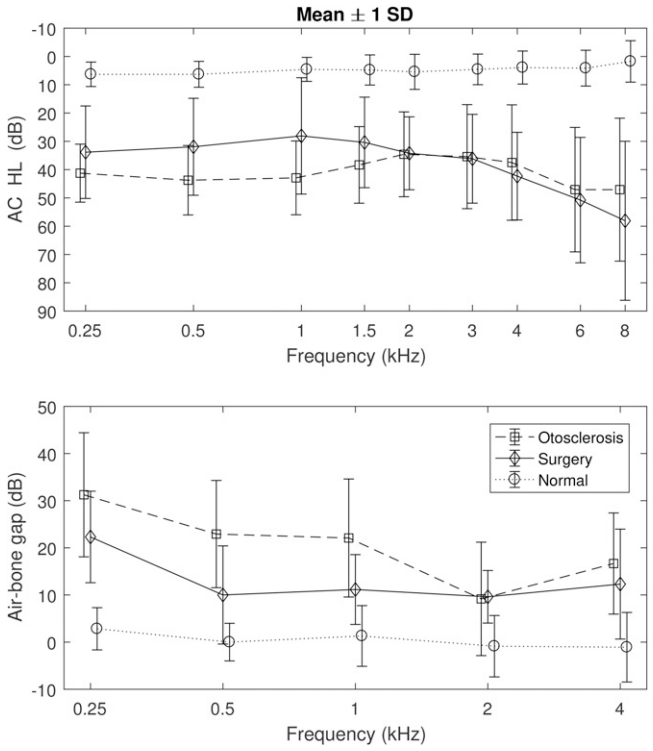

Figure 1. Mean ± 1 SD of AC audiograms (top) and air-bone gaps (bottom) are plotted for each group. Data are slightly adjusted along horizontal axis in this and later figures to improve clarity.

gap in the Surgery group was reduced by 9–14 dB from that in the Otosclerosis group between 0.25 and 1 kHz, which shows that the surgery to ameliorate conductive hearing loss achieved some success in the Surgery group. A sensorineural hearing loss was present in some ears in both the Surgery and Otosclerosis ears. This is consistent with the older age range (Table 1) in these groups relative to the Normal group and is also consistent with presence of otosclerosis, inasmuch as it may involve the cochlea as well as the stapes footplate, and the former can lead to a progressive sensorineural hearing loss.

TEOAE Results

Group results are presented for each individual research test, beginning with the TEOAE test inasmuch as the interpretation of its results is the simplest. Figure 2 shows the mean ± 1 SE of TEOAE SNR for each participant group at half-octave frequencies. The mean SNR in the Normal group was larger than the SNRs in the Otosclerosis and Surgery groups at all frequencies, whereas the mean SNRs of the Otosclerosis and Surgery groups were similar, with values less than the criterion SNR of 6.6 dB that is used to classify a TEOAE as present or absent at each frequency (Keefe et al, 2016). That is, the mean TEOAEs would be classified as absent in both the Otosclerosis and Surgery groups. The absent

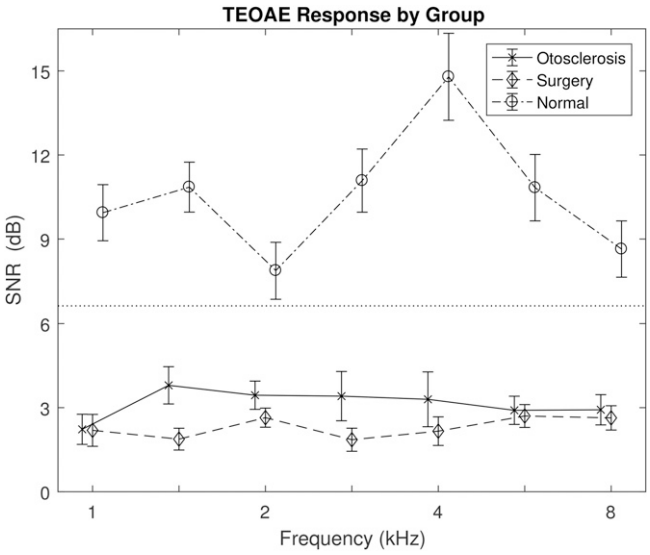

Figure 2. Mean ± 1 SE of TEOAE SNR responses are plotted. The horizontal dotted line at 6.6 dB is the minimum SNR at which a TEOAE would be classified as present (at $p = 0.05$).

TEOAEs in the Surgery group were likely due to a combination of a reduced level of reverse middle-ear transmission in ears receiving surgery to correct otosclerosis than in ears with normal middle-ear function, as well as to the possible presence of outer hair cell dysfunction in the Surgery group ears.

The LMM for TEOAE SNR showed significant main effects for group ($p < 0.0001$) and frequency ($p = 0.0358$) at each half-octave frequency between 1 and 8 kHz, and a significant interaction of group and frequency ($p = 0.0032$). The mean TEOAE SNR was significantly larger in the Normal group than the Surgery group at all frequencies (every $p < 0.001$) and significantly larger in the Normal group than the Otosclerosis group at all frequencies (every frequency had $p < 0.001$, except at 2 kHz for which $p = 0.009$).

Although not plotted, the LMM for TEOAE signal level showed significant main effects for group ($p < 0.0001$) and frequency ($p < 0.0001$) and a significant interaction of group and frequency ($p < 0.0001$). The mean TEOAE signal level in the Normal group was significantly larger than the Surgery group at all frequencies (all $p < 0.001$) and significantly larger than the Otosclerosis group at all frequencies (all $p < 0.001$, except $p = 0.003$ at 2 kHz). The mean signal levels in the Otosclerosis and Surgery groups had no significant differences across frequency.

The mean TEOAE noise levels showed significant main effects for group ($p < 0.0001$) and frequency ($p < 0.0001$), and a significant interaction of group and frequency ($p < 0.0001$). The mean TEOAE noise levels did not vary between groups at individual frequencies except for the following significant differences at low frequencies. The mean TEOAE noise levels at 1 and 1.4

kHz were 1–2 dB lower in the Otosclerosis group compared with the Normal group ($p = 0.004$ at 1 kHz, and $p = 0.025$ at 1.4 kHz) and the Surgery group ($p = 0.040$ at 1 kHz, and $p = 0.0240$ at 1.4 kHz). It is reasonable to consider that the measurement-system noise was similar in the three groups of test ears, so that the difference in noise was due to some physiological factor. One possible reason for this difference would be a slightly smaller contribution from blood flow-related noise to the noise measured in the ear canal in otosclerotic ears, although this topic requires additional research.

Clinical and Research ASR Results

Figure 3 shows the mean ± 1 SE of the clinical ASRT and research (or WB) ASRT as a function of ASR activator type for the Otosclerosis group (top panel), Surgery group (middle panel), and Normal group (lower panel). The clinical and research ASRT ranged up to 105 dB SPL for the six tonal activators and up to 80 dB SPL for the ipsilateral BBN activator. As described in "Methods" and shown on Figure 3, the NR response was encoded as 5 dB larger than the maximum activator

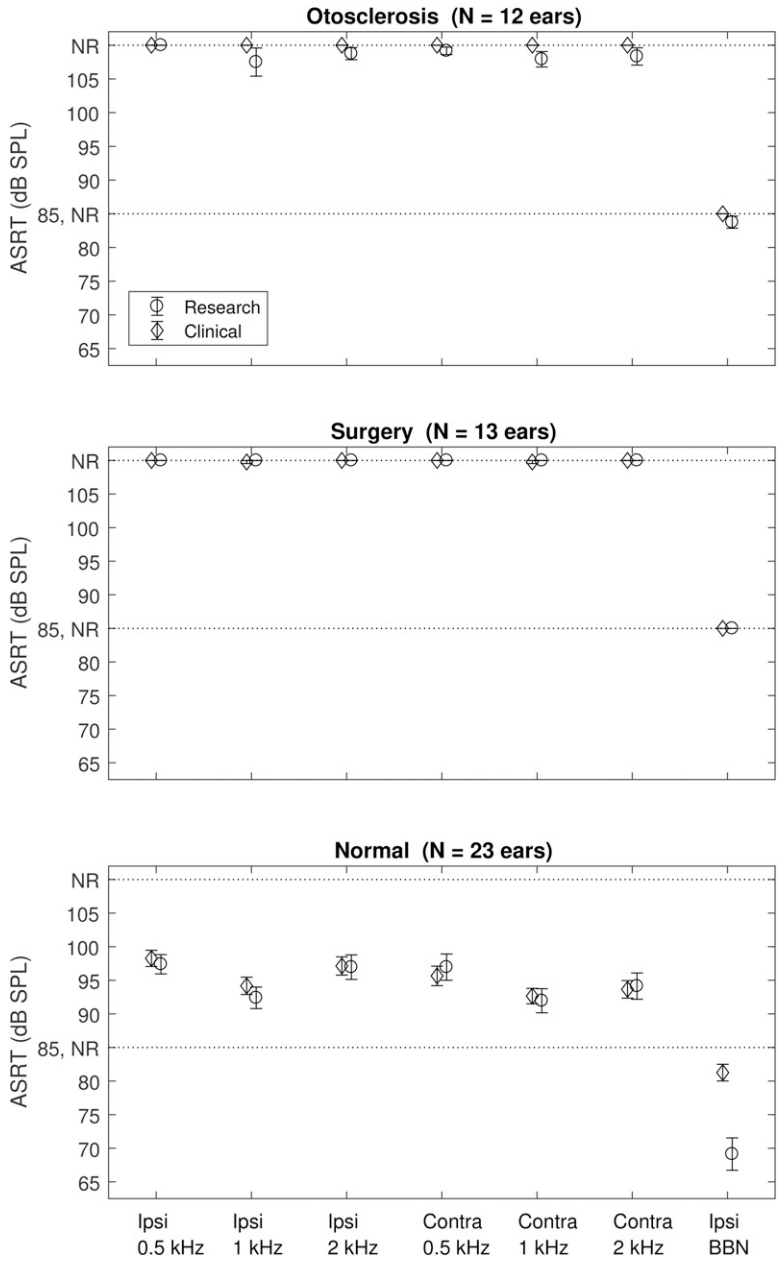

Figure 3. Mean ± 1 SE of ASRTs are plotted for the Otosclerosis group (top), Surgery group (middle), and Normal group (bottom) as a function of the type of ASRT test (Ipsi for ipsilateral test, and contra for contralateral test). Each panel shows ASRT data for the clinical test and the WB research test. On each panel, the upper dotted line shows NR data for tonal activators, and the lower dotted line shows NR data for the BBN activator.

level, i.e., as 110 dB SPL for tonal activators and 85 dB SPL for the ipsilateral BBN activator. The mean clinical and research ASRTs were at, or within 5 dB of, the NR level for all ASR types in both Otosclerosis and Surgery groups. Thus, the ASR was absent in most of these ears whether measured using the clinical or research test. Notwithstanding that fact, the mean ASRT trended to slightly lower levels for the research test than the clinical test, e.g., see Figure 3 (top panel) for the ipsilateral ASR test with the 1 kHz activator. In the Normal group (Figure 3, bottom panel), the mean ASRTs for tonal activators ranged from 92 to 99 dB SPL with high similarity between the clinical and research tests. As mentioned in the "Methods" section, the ASRTs for some of the activator types in the Surgery group were NRs for all ears, so that the accompanying SDs were zero. Because these ceiling effects influenced the assumption underlying the use of LMMs, this section places emphasis on the most clinically important results that included the largest mean differences in ASRT.

The main effects for group, ASRT-test-type, activator-ear, and activator for each activator-ear were significant (all $p < 0.001$). The following two-way interactions were significant: group and ASRT-test-type ($p = 0.0035$), ASRT-test-type and activator-ear ($p = 0.011$), ASRT-test-type and activator test for each activator-ear ($p < 0.001$), and group and activator test for each activator-ear ($p < 0.001$). The three-way interaction was significant between group, ASRT-test-type and activator test for each activator-ear ($p < 0.001$). This pattern of significance allowed the comparison of mean ASRTs at the combination of all factor levels.

The mean ASRTs in all three contralateral tests (Figure 3) were significantly lower for the Normal group than either of the other two groups (all $p < 0.001$). The mean ASRTs in all four ipsilateral tests (Figure 3) were significantly lower for the Normal group than the Surgery group (all $p < 0.001$), and lower for the Normal group than the Otosclerosis group over all activators, i.e., at 0.5 kHz ($p = 0.002$), 1 kHz ($p < 0.001$), 2 kHz ($p = 0.006$), and BBN ($p = 0.024$). There were no mean differences in ASRT between Otosclerosis and Surgery groups, inasmuch as the predominant value was an NR (Figure 3, top and middle panels).

The clinical and research ASRT tests were compared for the Normal group (Figure 3, bottom panel). The mean ASRT using the ipsilateral BBN activator was significantly lower by 13.6 dB for the research test compared with the clinical test ($p < 0.001$), with an effect size of 2.5. This large effect size may be because of the larger bandwidth of the BBN activator in the research test (up to 8 kHz) compared with the clinical test (up to about 4 kHz) under a condition of the same SPL in a 2-cm^3 coupler. There was no significant difference in the mean ASRTs for any of the tonal activator tests in the Normal group, that is, for ipsilateral and contralateral tonal activators at 0.5, 1, and 2 kHz (Figure 3, bottom panel).

Ambient Absorbance and Group Delay Results

Figure 4 (top panel) shows the mean ± 1 SE of the ambient absorbance (A_a) for each group. These mean differences were assessed for significance using an LMM, which showed main effects for group ($p = 0.0215$) and frequency ($p < 0.0001$), and a significant interaction of group and frequency ($p < 0.0001$).

The frequencies at which the mean difference in A_a between Surgery and Otosclerosis groups (S–O) was significant were 0.5 kHz ($p < 0.001$), 0.71 kHz ($p < 0.001$) and 1 kHz ($p = 0.042$). These significant differences are labeled on the top panel of Figure 4 with the text S–O. The frequencies at which the mean difference in A_a between Surgery and Normal groups (S–N) was significant were 0.5 kHz ($p < 0.001$), 0.71 kHz ($p < 0.001$), and 4 kHz ($p = 0.034$). These significant differences are labeled on the top panel of Figure 4 with the text S–N. The mean difference in A_a between Otosclerosis and Normal groups (O–N) was significant at 4 kHz ($p = 0.034$). No other differences in A_a were significant. This significant difference is labeled on the top panel of Figure 4 with the text O–N. Each significant mean difference in A_a had a large effect size (0.81–1.85). This convention of using S–O, S–N, and O–N to represent mean differences between pairs of groups in plotted results is used in subsequent plotted results described in the later paragraphs.

At 0.7 and 1 kHz, the mean absorbance trended to slightly smaller values in the Otosclerosis group than the Normal group, but the difference was not significant. This differs from previous research that showed a significantly smaller absorbance in otosclerotic ears than normal ears at frequencies from 0.4 to 1 kHz (Shahnaz, Bork, et al, 2009). A trend of increased absorbance in postoperative otosclerotic ears relative to normal ears at low frequencies has also been reported (Shahnaz, Longridge, et al, 2009). Although this difference was significant across all frequencies (i.e., the main effect of frequency was significant), the postoperative otosclerotic and normal ears did not have a significant difference at any particular frequency. By contrast, the present study found a significant difference in ambient absorbance between these two groups at 0.5, 0.7, and 4 kHz.

The mean ear-canal area and length were estimated in all test ears, and an LMM revealed that these ear-canal properties did not vary in the Normal, Otosclerosis, and Surgery groups. In the Normal, Otosclerosis, and Surgery groups, the mean ± 1 SE of the ear-canal lengths were 1.86 ± 0.12, 1.97 ± 0.12, and 2.06 ± 0.29 cm, with corresponding mean ± 1 SE round-trip

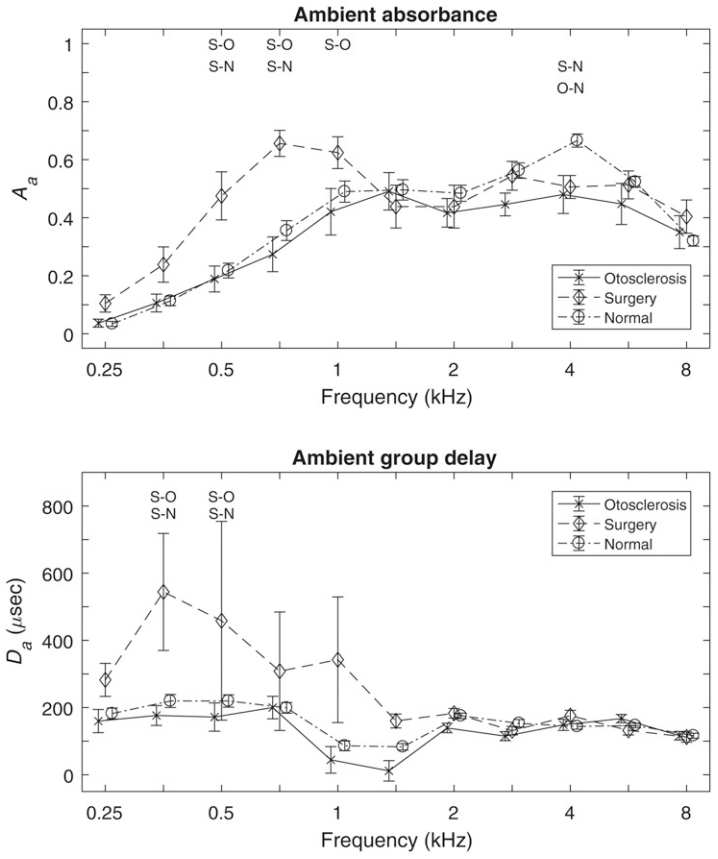

Figure 4. Mean ± 1 SE of ambient reflectance data are plotted for absorbance (top panel) and group delay (bottom panel). Each panel shows group data for Otosclerosis (O), Surgery (S), and Normal (N) ears with corresponding letter denoting each group. The pairs of letters denote pairs of group means that are significantly different at each frequency.

group delays (based on phase velocity 345 m/sec) of 108 ± 7, 119 ± 17, and 114 ± 12 μsec, respectively.

Figure 4 (bottom panel) shows the mean ± 1 SE of the ambient group delay (D_a) for each group. These mean differences were assessed for significance using an LMM, which showed a main effect of frequency (p = 0.0009). The mean D_a was significantly larger in the Surgery group than the Otosclerosis group at 0.35 kHz ($p < 0.001$) and 0.5 kHz ($p = 0.007$), and significantly larger in the Surgery group than the Normal group at 0.35 kHz ($p < 0.001$) and 0.5 kHz (p = 0.010). No other group mean differences in D_a were significant in Figure 4 (bottom panel). Each significant mean difference in D_a had a large effect size (0.81–1.69).

The group delay is approximately equal to the sum of the round-trip travel time between the probe tip and TM plus the group delay at the TM (Keefe et al, 2015). This round-trip travel time is twice the estimated length of the ear canal divided by the phase velocity of sound. Inasmuch as the mean estimates of ear-canal length were similar in normal and otosclerotic ears, any differences in group delay at the probe would be a measure of the difference in group delay at the TM. The mean ambient group delay at 8 kHz from Figure 4 (bottom) was 116, 112, and 116 μsec for the Normal, Otosclerosis, and Surgery groups, respectively, which are indeed similar to the mean round-trip group delays listed previously. A stiffened middle ear due to otosclerosis would be expected to have a reduced group delay, at least at low frequencies. Although the trend of the mean difference below 1 kHz supported this prediction, there were likely too few ears (Table 1) to adequately test for significance.

Tympanometric Absorbance and Group Delay Results

For each participant group, the mean tympanometric absorbance (A_t) is plotted in Figure 5 as a function of frequency at three air pressures in the downswept tympanogram: the positive-tail pressure (Ppt) (top panel), the TPP (middle panel), and the negative-tail pressure (Pnt) (bottom panel). The Ppt of +200 daPa was the most positive air pressure of the tympanometric sweep, and the Pnt of −300 daPa was the most negative air pressure. The results were similar between downswept and upswept tympanograms, so that only the results for the downswept tympanogram are described.

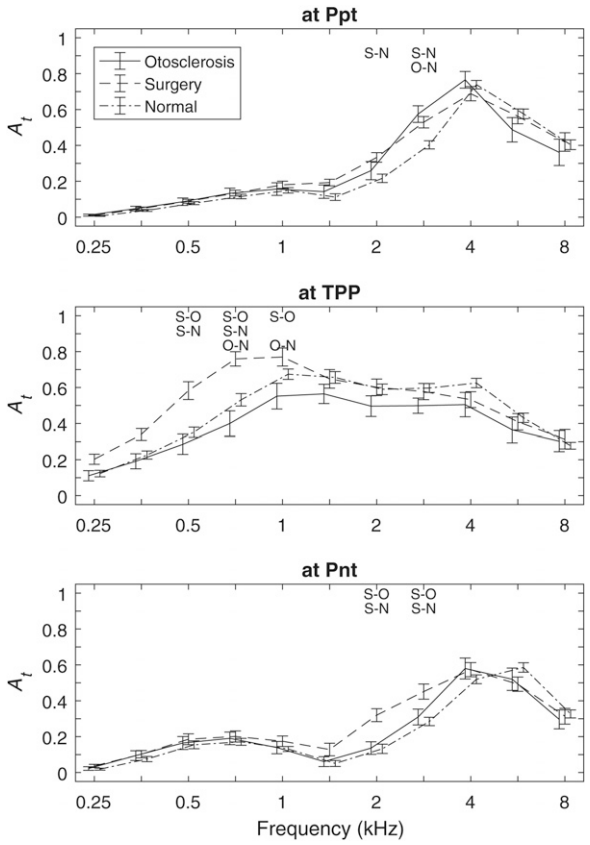

Figure 5. Mean ± 1 SE of tympanometric absorbance data are plotted at ear-canal air pressures of Ppt (top), TPP (middle), and Pnt (bottom). Each panel shows group data for Otosclerosis (O), Surgery (S), and Normal (N) ears with corresponding letter denoting each group. The pairs of letters denote pairs of group means that are significantly different at each frequency.

The absorbance A_t at TPP in the Normal group increased with increasing frequency up to 1 kHz, had relatively large values between 1 and 4 kHz, and decreased with increasing frequency between 4 and 8 kHz. This resembles the bandpass shape familiar from measurements of middle-ear transmission, although absorbance does not directly assess transmission from the ear canal to the cochlea. In the Normal group, the tympanometric absorbance A_t at TPP (Figure 5, middle) was similar to the ambient absorbance A_a in Figure 4 (top), inasmuch as the TPP in the normal group did not deviate much from ambient pressure. In the Normal group, the A_t at TPP was larger than A_t at either tail pressure in the Normal group at all frequencies between 0.25 and 2 kHz. This peak-to-tail difference in the tympanometric absorbance in normal ears is described in more detail in Feeney et al (2016). The middle ear absorbs more energy at the TPP than at the tail pressures, although the absorbance does exceed zero at the tail pressures. The latter contradicts an assumption underlying the compensation procedure for 226-Hz admittance tympanometry (Keefe et al, 2015).

An additional factor of pressure with factor levels (Ppt, TPP, and Pnt) was included in the LMMs for tympanometric absorbance and tympanometric group delay. This factor incorporated the variation of tympanometric reflectance variables across the range of air pressures used in the tympanogram at TPP relative to each of the tail pressures.

The mean differences in A_t were assessed for significance using an LMM, which showed main effects for pressure ($p < 0.0001$), group ($p = 0.0088$) and frequency ($p < 0.0001$), and significant two-way interactions of pressure and group ($p = 0.0002$) and pressure and frequency ($p < 0.0001$). The three-way interaction of pressure, group, and frequency was significant ($p < 0.0001$). The three-way interaction of A_t was evaluated with group differences for each pressure value (Ppt, TPP, and Pnt) and each half-octave frequency.

For tympanometric absorbance at TPP (Figure 5, middle panel), the low-frequency pattern of mean differences in A_t at TPP resembled that for A_a. The mean A_t at TPP was larger in the Surgery group than the Normal group at 0.5 kHz ($p < 0.001$) and 0.71 kHz ($p < 0.001$), and larger in the Surgery group than the Otosclerosis group at 0.5 kHz ($p < 0.001$), 0.7 kHz ($p < 0.001$), and 1 kHz ($p = 0.002$). The mean A_t at TPP was smaller in the Otosclerosis group than the Normal group at 0.71 kHz ($p = 0.016$) and 1 kHz ($p = 0.049$). All effect sizes were large (0.81–2.27) for significant mean differences in A_t at TPP.

The mean difference in A_t at Ppt (Figure 5, top panel) was larger in the Surgery group than the Normal group at 2 kHz ($p = 0.015$) and 2.8 kHz ($p = 0.005$). The mean A_t at Ppt was larger in the Otosclerosis group than the Normal group at 2.8 kHz ($p < 0.001$). All effect sizes were large (0.99–1.45) for significant mean differences in A_t at Ppt.

The mean difference in A_t at Pnt (Figure 5, bottom panel) was larger in the Surgery group than the Normal group at 2 kHz ($p < 0.001$) and 2.8 kHz ($p < 0.001$), and larger in the Surgery group than the Otosclerosis group at 2 kHz ($p < 0.001$) and 2.8 kHz ($p = 0.006$). All effect sizes were large (1.18–1.56) for significant mean differences in A_t at Pnt.

Pressure asymmetries in absorbance were evident above 0.7 kHz with larger A_t in the Otosclerosis and Surgery groups at 2–4 kHz in the Ppt response compared with the corresponding group means in the Pnt response (Figure 5, top and bottom panels). These across-frequency effects on mean A_t derived from the significant interaction on pressure and group. Moreover, the mean A_t at Ppt was significantly larger in the Otosclerosis group than the Normal group at 2.8 kHz, whereas the mean A_t at Pnt was not. Such differences in tympanometric absorbance at the tail pressures have no comparative condition in the ambient absorbance measurement. The Ppt and Pnt conditions

result in inward and outward relative forces applied to the TM, respectively, which generate inward and outward displacements of the ossicular chain via the manubrium that is in contact with the interior surface of the TM. This suggests that the tail asymmetry in absorbance may be of diagnostic value in identifying otosclerotic ears.

For each participant group, the mean tympanometric group delay (D_t) is plotted in Figure 6 as a function of frequency at three air pressures in the downswept tympanogram: the Ppt (top panel), the TPP (middle panel), and the Pnt (bottom panel). The tympanometric group delays in the Normal group were larger at TPP than at the tail pressures at frequencies from 0.25 to 0.7 kHz, and smaller at 1 kHz. The mean D_t had much less variation with frequency at the tail pressures than at TPP.

The mean D_t had significant two-way interactions between pressure and frequency ($p < 0.0001$), and group and frequency ($p = 0.022$). The mean group differences in D_t at TPP (Figure 6, middle) were significant only at 0.35 kHz, for which: D_t was larger in the Surgery group than the Normal group ($p = 0.041$), larger in the Surgery group than the Otosclerosis group ($p < 0.001$), and larger in the Normal group than the Otosclerosis group ($p = 0.047$). This latter result contrasts with the mean D_a that did not differ in the Normal and Otosclerosis groups at any frequency (see lower panel, Figure 4). The larger values of group delay in the Surgery group at 0.035 kHz may be related to functional effects of the middle-ear prosthesis that are present in the Surgery group ears. All effect sizes of significant mean differences in D_t at TPP were large (1.99–3.80).

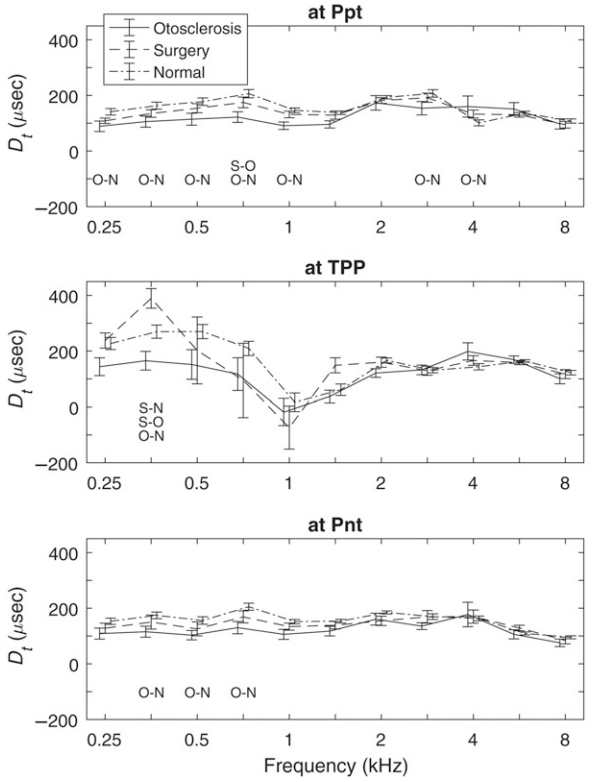

Figure 6. Mean ± 1 SE of tympanometric group delay data are plotted at ear-canal air pressures of Ppt (top), TPP (middle), and Pnt (bottom). Each panel shows group data for Otosclerosis (O), Surgery (S), and Normal (N) ears with corresponding letter denoting each group. The pairs of letters denote pairs of group means that are significantly different at each frequency.

Mean group differences in D_t at the tail pressures were evident in Figure 6 over a wider frequency range than at TPP, with maximum differences between conditions observed at frequencies up to and including 4 kHz. The mean D_t at Ppt (Figure 6, top) was smaller in Otosclerosis than Normal ears at lower frequencies, i.e., at 0.25 kHz ($p = 0.038$), 0.35 kHz ($p = 0.018$), 0.5 kHz ($p = 0.008$), 0.71 kHz ($p < 0.001$), and 1 kHz ($p = 0.025$). The mean D_t at Ppt was also smaller in Otosclerosis than Normal ears at 2.8 kHz ($p = 0.033$), but larger at 4 kHz ($p = 0.013$). The mean D_t at Ppt was smaller in Otosclerosis than Surgery ears at 0.71 kHz ($p = 0.049$). All effect sizes of significant mean differences in D_t at Ppt were large (2.26–4.05).

The mean D_t at Pnt (Figure 6, bottom) was also smaller in Otosclerosis ears than Normal ears, but over a smaller low-frequency range at 0.35 kHz ($p = 0.019$), 0.5 kHz ($p = 0.043$), and 0.7 kHz ($p = 0.001$). This reduced D_t at the tail pressures at low frequencies is further evidence that the middle-ear stiffness was increased in otosclerotic ears, so that less sound energy was absorbed from the click stimulus at these frequencies in the tail-pressure conditions. Pressure asymmetries in tympanometric group delay at Ppt and Pnt were evident in otosclerotic ears at lower frequencies than those observed for tympanometric absorbance. This is evidence of the potential value of tympanometric group delay measurements in classifying ears with otosclerosis. All effect sizes were large (2.46–3.51) for significant mean differences in D_t at Pnt.

Overall, the D_t at Ppt test revealed mean differences between Otosclerosis and Normal ears at more frequencies than the corresponding D_t responses at TPP and Pnt. This is evidence of a potential diagnostic advantage at using the tympanometric reflectance rather than the ambient reflectance test in differentiating Normal from Otosclerotic ears. The effect sizes were larger for tympanometric group delay than for other tests (tympanometric absorbance, ambient absorbance, and ambient group delay).

Figure 7 shows the mean ± 1 SE of the differences in absorbance ΔA_t among the three pairs of tympanometric measurements at Ppt, TPP, and Pnt. The top panel shows the peak-to-positive-tail difference ΔA_t of A_t (TPP) minus A_t (Ppt), and the middle panel shows the peak-to-negative-tail difference ΔA_t of A_t (TPP) minus A_t (Pnt). In the Normal group, the peak-to-tail difference for each tail was positive at low frequencies with

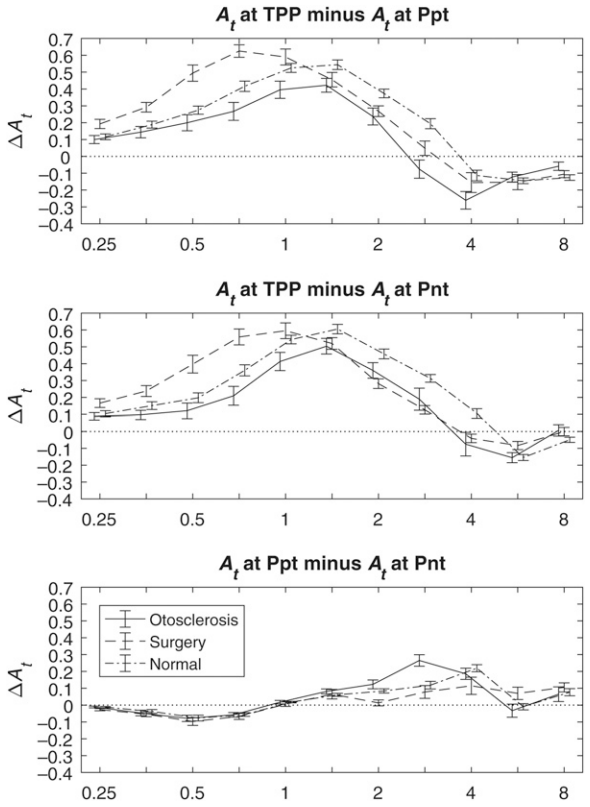

Figure 7. Mean ± 1 of the difference in tympanometric absorbances ΔA_t are plotted at ear-canal air pressures of Ppt (top), TPP (middle), and Pnt (bottom) for each participant group. Top: ΔA_t for A_t (TPP) − A_t (Ppt). Middle: ΔA_t for A_t (TPP) − A_t (Pnt). Bottom: ΔA_t for A_t (Ppt) − A_t (Pnt).

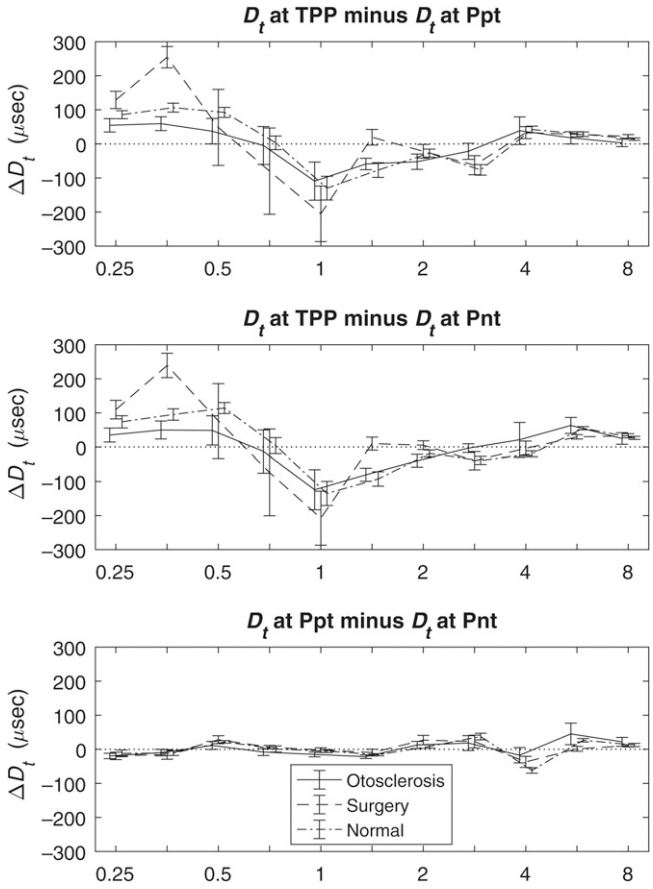

Figure 8. Mean ± 1 of the difference in tympanometric group delays ΔD_t are plotted at ear-canal air pressures of Ppt (top), TPP (middle), and Pnt (bottom) for each participant group. Top: ΔD_t for D_t (TPP) − D_t (Ppt). Middle: ΔD_t for D_t (TPP) − D_t (Pnt). Bottom: ΔD_t for D_t (Ppt) − D_t (Pnt).

a local maximum at 1.4 kHz, and negative above 4–5.7 kHz. Comparing trends in the top and middle panels, the Otosclerosis group had a smaller maximum peak-to-tail difference in mean absorbance compared with the Normal group at frequencies up to and including 4 kHz. The pressure asymmetry in A_t was assessed in the bottom panel as the absorbance difference at the positive tail relative to that at the negative tail, i.e., a ΔA_t of A_t (Ppt) minus A_t (Pnt). This pressure asymmetry difference had a maximum in the Otosclerosis group at 2.8 kHz compared with lesser maxima in the Normal and Surgery groups at 4 kHz. This asymmetry of the absorbance tympanogram at these relatively high frequencies may be related to a difference in transmission through the ossicular pathway in otosclerotic ears relative to the other ears.

Figure 8 shows the mean ± 1 SE of the differences in group delay ΔD_t among the three pairs of tympanometric measurements at Ppt, TPP, and Pnt. The top panel shows the peak-to-positive-tail difference ΔD_t of D_t (TPP) minus D_t (Ppt), and the middle panel shows the peak-to-negative-tail difference ΔD_t of D_t (TPP) minus D_t (Pnt). In the Normal group, the peak-to-tail difference for each tail was on the order of 100 μsec at frequencies near 0.35–0.5 kHz, and on the order of −100 μsec at 1 kHz, with smaller magnitudes at higher frequencies. The pressure asymmetry in D_t is assessed in the bottom panel as the group delay difference at the positive tail relative to that at the negative tail, i.e., a ΔD_t of D_t (Ppt) minus D_t (Pnt). This mean difference was smaller in magnitude with more variability near 4–5.7 kHz. Comparing trends in the two peak-to-tail differences in mean group delay across the three groups (Figure 8, top and middle), the Otosclerosis group had smaller magnitudes between 0.25 and 1 kHz than did the Normal group, and both had similar fine structure. The bottom panel shows the pressure asymmetry difference in group delay, for which the group differences were small.

These measurements of peak-to-tail mean differences in ΔA_t and ΔD_t for the Normal group replicate findings for normal adult ears (Feeney et al, 2016). The peak-to-tail mean differences in ΔA_t and ΔD_t for the Surgery group at frequencies below 1 kHz (top and middle panels of Figures 7 and 8, respectively) were of larger magnitudes than for the other groups, which were similar in

pattern to the mean A_t and D_t at TPP for the Surgery group in the middle panels of Figures 5 and 6, respectively. An interesting finding is that the significant mean difference in the O–N comparison in A_t at Ppt at 2.8 kHz (Figure 5, top) appears to coincide with the large pressure asymmetry in ΔA_t in the Otosclerosis group at the same frequency (Figure 7, bottom).

Univariate Reflectance Classifiers for Otosclerosis and Normal Groups

The preceding analyses identified which mean differences in reflectance test variables were significant among the Normal, Otosclerosis, and Surgery groups of ears. A clinically important task for diagnosing otosclerosis is to classify a particular test ear as normal or otosclerotic, as inclusion in a postsurgery group is explicitly known. In this respect, the LMM results are suggestive, inasmuch as any reflectance variable that had different means in the Normal and Otosclerosis groups would be a candidate variable to use in a diagnostic test. For the ambient absorbance and group delay test (Figure 4), the only significant mean difference between Normal and Otosclerosis groups was A_a at 4 kHz. For the tympanometric absorbance and group delay test results in Figures 5 and 6, the only significant mean differences between Normal and Otosclerosis groups were A_t at TPP at 0.7–1 kHz, A_t at Ppt at 2.8 kHz, D_t at Ppt at 0.25–1 kHz and 2.8–4 kHz, and D_t at Pnt at 0.35–0.7 kHz. Even if a particular mean difference was highly significant, the distributions of means overlapped between the test scores in the Normal and Otosclerosis groups.

The AUC for each single ambient-reflectance variable to classify ears in the Normal and Otosclerosis groups was calculated from the measured data without regard to these LMM results. Figure 9 (top panel) shows plots of these AUCs for the ambient absorbance A_a and ambient group delay D_a. The AUCs varied from 0.5, which corresponds to no diagnostic accuracy, to 0.80, which corresponds to the best relative accuracy, although less than the ideal accuracy of 1. The rank order of the six

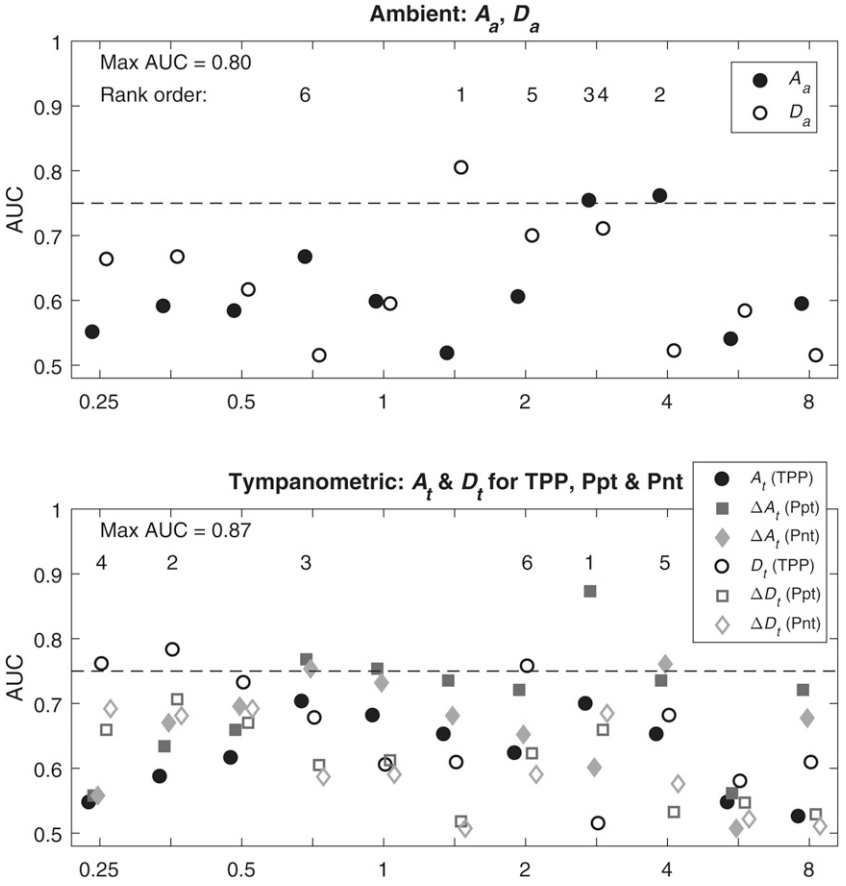

Figure 9. Top: AUCs for individual ambient reflectance test variables A_a (filled circle markers) and D_a (open circle markers) as functions of frequency. Bottom: AUCs for individual tympanometric reflectance test variables A_t and ΔA_t (filled markers) and D_t and ΔD_t (open markers) as functions of frequency. A_t and D_t are measured at TPP (circle marker). ΔA_t (Ppt) is A_t(TPP) − A_t(Ppt), ΔA_t (Pnt) is A_t(TPP) − A_t(Pnt). ΔD_t (Ppt) is D_t (TPP) − D_t (Ppt), and ΔD_t (Pnt) is D_t (TPP) − D_t (Pnt). The variables at Ppt use square markers and those at Pnt use diamond markers. In both panels, a horizontal dashed line is drawn through an AUC of 0.75 as a visual aid, the rank order of the six largest AUCs is listed across the top of the panel, and the maximum AUC across all variables is specified in the upper left corner.

largest or best AUCs is listed across the top of the panel, and the maximum AUC across all variables is specified in the upper left corner. These six variables included responses between 0.7 and 4 kHz. The best single predictor was the group delay at 1.4 kHz, which was not significant in the corresponding LMM. The second best predictor was the ambient absorbance at 4 kHz, which was significant in the LMM. The third best predictor was ambient absorbance at 2.8 kHz, which was also not significant. Previous studies described previously have identified the ambient absorbance in the range from 0.4 to 1 kHz as the best predictor for otosclerosis, but the only such variable in the best six AUCs was the ambient absorbance at 0.7 kHz (Figure 9, top).

Based on the group measurements of peak-to-positive-tail and peak-to-negative tail differences of ΔA_t and ΔD_t (Figures 7 and 8, top and middle panels), the tympanometric variables to classify ears in the Normal and Otosclerosis groups were selected to be A_t and D_t at TPP, ΔA_t for the pair of peak-to-tail absorbance differences, and ΔD_t for the pair of peak-to-tail group delay differences. Figure 9 (bottom panel) shows plots of these AUCs for the tympanometric data. The maximum AUC was 0.87, which was larger than any AUC for the ambient reflectance test. The rank order of the six largest or best AUCs is listed across the top of the panel, and the maximum AUC across all variables is specified in the upper left corner. These six variables included responses between 0.25 and 4 kHz. The best predictor was ΔA_t at Ppt at 2.8 kHz. This was related to A_t at Ppt at 2.8 kHz, which was a significant variable in the LMM (Figure 5, top). The second best predictor was D_t at TPP at 0.35 kHz, and the third best was ΔA_t at Ppt at 0.70 kHz. This latter variable incorporated A_t at TPP, which was significant at 0.70 kHz in the LMM (Figure 5, middle). Nevertheless, the A_t at TPP variable at 0.70 kHz was not among the six variables with the largest AUCs (Figure 9, bottom)

ROC Curves for Reflectance

This subsection has quantitative details on reflectance test performance to classify ears as in the Normal or Otosclerosis groups. The ROC curves are shown for the ambient reflectance and tympanometric reflectance variables having the largest AUCs (Figure 10). A test criterion was selected for which the distance was a minimum between the ROC curve value and the point representing perfect test performance (i.e., sensitivity and specificity equal to 1), which would lie at the upper left hand corner of the ROC curve. These test criterion values are indicated in Figure 10 by a diamond for the ambient reflectance test and a circle for the tympanometric reflectance test.

This definition of test criterion is related to the point of symmetry on the ROC curve for the case of a predictor

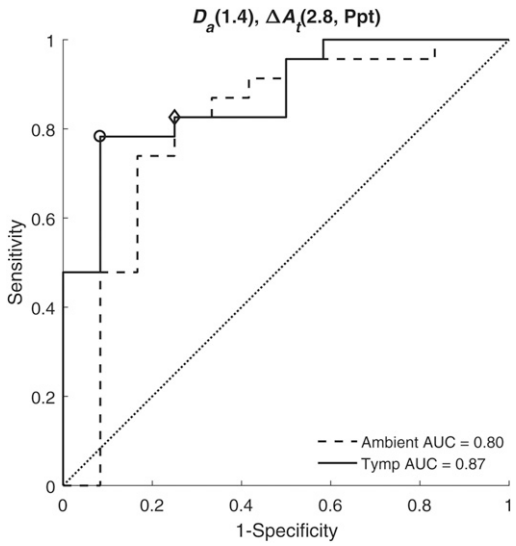

Figure 10. ROC curves are plotted for the ambient and tympanometric test with the largest AUC. The legend shows the AUC for each test for the test variable listed in the title of the panel (D_a at 1.4 kHz, ΔA_t at 2.8 kHz at Ppt). A diamond marker on the ROC curve of the ambient test and a circle marker for the tympanometric test list the point on the ROC curve with minimum distance from ideal test performance.

with continuous values. The point of symmetry is the point on the ROC curve with equal specificity and sensitivity, which essentially gives equal weighting to reducing the cost of the false positives and false negatives. In each case, the test criterion in Figure 10 is a point on the ROC curve with large specificity and large sensitivity.

Table 2 lists the three largest ambient reflectance variables by AUC and the three largest tympanometric reflectance variables by AUC. For each test variable, the criterion value to classify a test ear as Normal or Otosclerosis is listed along with its criterion specificity, criterion sensitivity, and criterion distance from ideal performance. These criterion values may be interpreted in relation to the ambient reflectance means in Figure 4 and the tympanometric reflectance means in Figures 5–8 as to whether larger values than the criterion corresponded to the Normal or Otosclerosis group. For example, a test value of D_a at 1.4 kHz $\geq$45.9 μsec was classified as Normal (see row 1 of Table 2 and Figure 4).

Multivariate Reflectance Classifiers for Otosclerosis and Normal Groups

It appeared inefficient to discard all data from a WB test except for a single variable that performed best across all univariate tests. The number of analyzed variables was 22 in the ambient reflectance test, and 66 in the tympanometric reflectance test (Figure 9). Yet the numbers of test ears was only 23 in the Normal group and 12 in the Otosclerosis group (Table 1), which were

Table 2. ROC Analyses of Reflectance Test Variables to Classify Ears as Normal or Otosclerosis

Test Type	Rank Order	Variable	Criterion Value	Criterion Specificity	Criterion Sensitivity	Criterion Distance	Mean	SD
Ambient	1	D_a (1.4 kHz)	45.9 μsec	0.750	0.826	0.305	58.6 μsec	78.3 μsec
	2	A_a (4 kHz)	0.559	0.583	0.913	0.426	0.602	0.180
	3	A_a (2.8 kHz)	0.500	0.750	0.739	0.362	0.522	0.140
Tymp.	1	ΔA_t (2.8 kHz, Ppt)	0.120	0.917	0.783	0.233	0.102	0.204
	2	D_t (0.35 kHz, TPP)	185 μsec	0.750	0.783	0.331	233 μsec	124 μsec
	3	ΔA_t (0.7 kHz, Ppt)	0.292	0.667	0.783	0.398	0.365	0.174

Note: Those variables with the three largest AUCs are listed in rank order.

smaller than the 66 variables in the tympanometric test. A goal was to use more of the WB test data in a multivariate classifier of ear status, while not using so many test variables that the results would be unlikely to generalize to a new population. A significant property of the ambient absorbance test is that its responses are highly correlated across frequency (Werner et al, 2010). The same is likely true for the ambient group delay and the tympanometric absorbance and group delay tests.

The number of multivariate test variables may be reduced using a principal component analysis (PCA), which results in multivariate outputs that are uncorrelated with one another. Because PCA outputs are sorted in terms of the amount of total variance accounted, those outputs can be discarded that explain only a small amount of variance. The PCA identifies a smaller dimensional subspace of the multivariate responses spanned by a set of uncorrelated test vectors called the principal components (PCs). However, the PCA acts over both normal and impaired groups, so that the components obtained may or may not be helpful in the classification task. That is, a component that explains much of the total variance across both groups may be inaccurate in classifying a given test into its true group.

An alternative ROC-supervised PCA approach was devised to classify mice as normal or having Huntington disease (Nikas and Low, 2011). The AUC was first calculated for each input test variable in a univariate test for its ability to correctly classify each response into its true normal or impaired group, as was described in the present study in the preceding section on univariate classifiers for otosclerosis. Only those test variables were retained for which AUC exceeded a critical AUC value and these retained components were inputs to a PCA. This supervision restricted the PCA to a subset of variables that were the best univariate classifiers. The AUC for each output component from the PCA was next calculated, and those components with AUC less than the criterion were removed to form a smaller set of test variables. They iterated on this approach with ever-increasing values of the critical AUC and used the final component that explained the most variance as the classifier. Their results showed promise, but appeared to suggest that the component that explained the most variance may not have been the best classifier.

Differences in the present study from that of Nikas and Low are that the datasets are smaller so that it was not possible to separate each group into a test and validation dataset, and the total number of test variables was much larger. Thus, a new multivariate classifier procedure was devised for the present study that was efficient at excluding test variables, yet defined statistically independent variables in the classifier. Unlike Nikas and Low, only a single PCA was calculated for each selected number of test variables.

ROC-Supervised PCA Results

A ROC-supervised PCA was first performed on the rank-order list of the M ambient reflectance test variables with the largest AUCs and the M tympanometric reflectance test variables with the largest AUCs. The multivariate M was varied from two up to six. Each reflectance test variable was first converted to a standardized z score before calculating the PCA. This conversion corrected for the fact that the physical units and data ranges varied considerably across the absorbance and group delay variables. An arbitrary test variable x measured across all ears in both of the Normal and Otosclerosis groups has a sample mean m_x and SD SD_x. The standardized z score for the i-th test ear with value x_i is $z_i=(x_i - m_x)/SD_x$. The mean and SD of the measured test variables with the three largest AUCs are listed in Table 2 for the ambient and tympanometric reflectance tests.

Summing across Normal and Otosclerosis groups, there were reflectance test scores for N participants and M variables in the PCA of order M. Each z score was represented by a matrix $\{z_{ij}\}$ with i labeling the N rows and j the M columns. The PCA was calculated using singular value decomposition using coefficients weighted by the inverse sample variance. This means that the PCs were based on the correlation matrix of the test data, which reduced the effect of variables with larger sample variance. One PCA output was the M by M matrix $\{P_{ij}\}$ of PCs, in which each column of the

matrix held a PC and each row had the loading for the M test variables. The PCs were listed in order of the decreasing amounts of variance accounted for in the input matrix $\{z_{ij}\}$. The first PC explained the most variance in the M reflectance test variables.

Another PCA output was the N by M matrix of PC scores $\{s_{ij}\}$, which represented the original reflectance data in the PC space. The matrix $\{s_{ij}\}$ was calculated as the matrix product of $\{z_{ij}\}$ right multiplied by $\{P_{ij}\}$. The PC matrices $\{P_{ij}\}$ are listed in Table 3 for $M = 2$ and $M = 3$ for both the ambient and tympanometric reflectance. These can be used to calculate PC scores in a new set of reflectance data by: (1) choosing M of 2 or 3, (2) selecting the particular reflectance variables from Table 2, (3) converting these corresponding new reflectance test values to z scores using the mean and SD of the reflectance data in Table 2, and (4) right multiplying the new data $\{z_{ij}\}$ by $\{P_{ij}\}$ in Table 3 to obtain $\{s_{ij}\}$. These steps can be followed using data with a single test ear ($N = 1$) or multiple test ears ($N > 1$ rows).

Logistic Regression Results

To address the difficulty that the component explaining the most variance may not be the best predictor, a final step was added in which the PC scores from the M best performing PCs were combined using a logistic regression procedure (Hosmer and Lemeshow, 2000) to output a univariate classifier into Normal or Otosclerosis groups. The diagnostic accuracy of this output classifier was assessed in terms of its AUC.

The fact that each component from the PCA was uncorrelated with any other component solved the problem of correlated inputs in a multivariate classifier, which would otherwise degrade its performance. The number of multivariate input variables (M) was varied from one to six in the analyzed results for each test. These six ambient reflectance variables are in the rank-ordered list in Figure 9 (top), and the six tympanometric variables are in the rank-ordered list in Figure 9 (bottom).

Table 3. PCA Results for Reflectance Test Variables Combined Across Normal and Otosclerosis Groups for those Two Variables with the Largest AUCs (Top) and those Three Variables with the Largest AUCs (Bottom)

$M = 2$ table of PCA results					
Test Type	PC (i)	Variance Explained (%)	$\{P_{i1}\}$	$\{P_{i2}\}$	
Ambient	1	59.6	0.7071	−0.7071	
	2	40.4	0.7071	0.7071	
Tymp.	1	69.7	0.7071	−0.7071	
	2	30.3	0.7071	0.7071	
$M = 3$ table of PCA results					
Test Type	PC (i)	Variance Explained (%)	$\{P_{i1}\}$	$\{P_{i2}\}$	$\{P_{i3}\}$
Ambient	1	53.1	0.3307	0.9346	−0.131
	2	31.2	0.6807	−0.1401	0.719
	3	15.7	0.6536	−0.3269	−0.6826
Tymp.	1	69.4	0.3987	0.8955	−0.1977
	2	20.4	0.671	−0.1379	0.7285
	3	10.2	0.6251	−0.4232	−0.6559

For each M, the logistic regression procedure calculated the coefficients β_j for the j-th PC and a constant coefficient β_0 (that did not affect test performance in terms of AUC). The multivariate classifier y_i of the i-th ear test was calculated by $y_i = \beta_0 + \sum_{j=1}^{M} s_{ij}\beta_j$ in terms of the PCA scores s_{ij}. The quantity e^{β_j} is the odds ratio of the j-th PC, that is, the amount that the j-th PC contributes to the classifier. The probability π_i of otosclerosis for the i-th ear is calculated as $\pi_i = e^{y_i}/(1 + e^{y_i})$.

The logistic regression coefficients for $M = 2$ were $\beta_0 = 0.7090$, $\beta_1 = 2.3974$, and $\beta_2 = 0.0085$ for the ambient test, and $\beta_0 = 1.4354$, $\beta_1 = 2.3887$, and $\beta_2 = 0.6704$ for the tympanometric test. The fact that β_2 was close to zero for the ambient test means that its multivariate classifier was strongly weighted on the first PC. By contrast, the multivariate classifier for the tympanometric test data used information from both PCs. The coefficients for $M = 3$ were $\beta_0 = 0.8439$, $\beta_1 = 2.2739$, $\beta_2 = 1.2500$, and $\beta_3 = 0.167$ for the ambient test, and $\beta_0 = 2.8553$, $\beta_1 = 3.7832$, $\beta_2 = 2.9208$, and $\beta_3 = 3.5684$ for the tympanometric test.

The test performance of each multivariate classifier on the Normal and Otosclerosis groups was assessed in terms of the ROC curve, from which its corresponding AUC was calculated. Figure 11 shows these AUCs for the ambient and tympanometric reflectance tests. The AUCs for $M = 1$ are the same as the maximum AUCs in Figure 9 for each test (because no PCA was performed for $M = 1$). The AUC increased with increasing M. For any given M, the tympanometric reflectance test was more accurate than the ambient reflectance test. Because there were only 12 otosclerotic ears in the study, a value of $M = 3$ (much less than 12) was selected for the final results, for which AUC was 0.95 for the tympanometric test and 0.88 for the ambient test. There was a large increase in AUC for either test in the $M = 2$ classifier compared with the $M = 1$ classifier, and both showed a further increase for $M = 3$. As is evident from the rank order lists in Figure 9 and Table 2, increasing M from 1 to 3 increased the frequency range of the variables in the classifier as well as the number of variable types, i.e., including absorbance and group delay variables, and tympanometric variables at both TPP and Ppt.

The criterion value of the multivariate classifier was calculated as before based on the minimum distance on the ROC curve from ideal performance. For the ambient test, the critical value of the probability of otosclerosis was 0.2746 for $M = 2$ and 0.5824 for $M = 3$. For the tympanometric test, the critical value was 0.2948 for $M = 2$ and 0.2392 for $M = 3$. These can be used to calculate multivariate classifier scores in a new set of reflectance data by

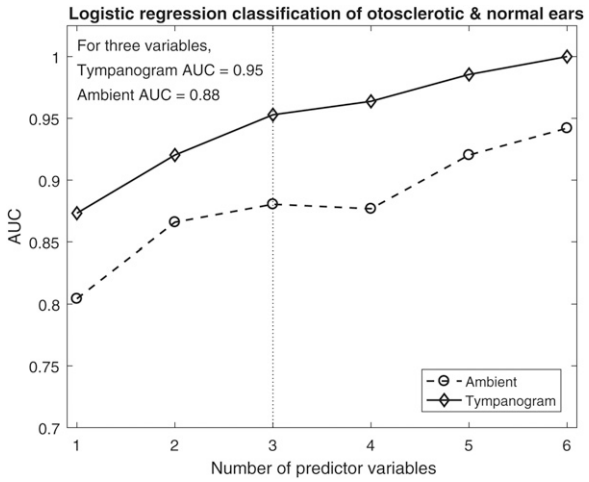

Figure 11. AUCs are plotted for the multivariate classifiers based on the ambient reflectance and tympanometric reflectance test responses as a function of the number (M) of variables in the classifier.

adding to steps 1–4 aforementioned as follows: (5) calculate multivariate classifier scores y_i and π_i in terms of the PCA scores $\{s_{ij}\}$ and the values of β_0 and β_j listed above, and (6) classify the ear as otosclerotic if π_i is greater than or equal to the critical value of the probability of otosclerosis, or otherwise classify the ear as normal.

DISCUSSION

Novel aspects of this otosclerosis study included the analyses of group delay data measured at ambient pressure in the ear canal, and analyses of tympanometric absorbance and group delay data. The mean ambient absorbance data were not significantly different at low frequencies up to 1 kHz in the Normal and Otosclerosis groups, which contrasted with previous results in a larger group of ears (Shahnaz, Bork, et al, 2009), although both studies are in the direction of reduced absorbance at low frequencies in otosclerotic ears. These low-frequency effects may be due to a stiffening of the TM due to the relative stiffening of the ossicular chain, which loads the TM. The reduced ambient absorbance in the Otosclerosis group at 4 kHz was significant and represents a novel finding, and may be due to functional differences in ossicular-chain transmission at frequencies above the dominant resonance of the TM. Mean differences were observed in tympanometric absorbance at TPP at low frequencies (0.7–1 kHz) and in the absorbance at the positive-tail pressure (at 2.8 kHz). The mean group delay was significantly lower in both tail pressure measurements at lower frequencies, and differed between Otosclerosis and Normal groups at the positive-tail pressure at higher frequencies, with reduced group delay in the Otosclerosis group at 2.8 kHz, and larger group delay at 4 kHz.

The Surgery group had larger ambient absorbance than either of the diagnosed Otosclerosis and Normal groups at lower frequencies (0.5 to 0.7–1 kHz), and lower absorbance than the Normal group at 4 kHz. The Surgery group had larger ambient group delay than the other groups at low frequencies (0.35–0.5 kHz). It had larger tympanometric absorbance at TPP at lower frequencies (0.5 to 0.7–1 kHz) and larger absorbance at the tail pressures than the Normal group at 2.8–4 kHz. The tympanometric group delay at TPP in the Surgery group also had a distinct pattern from Normal and Otosclerosis groups at 0.35–0.5 and 1 kHz, but showed no differences at the tail pressures. The additional information in the tympanometric absorbance and group delay appeared to have diagnostic relevance between groups compared with that in the corresponding ambient-pressure tests.

It is evident in Figure 1 (bottom) that the air-bone audiometric gaps in Surgery ears were only partially ameliorated on average by the introduction of the middle-ear prosthesis compared with the air-bone gaps in Normal ears. This is evidence that the surgery to improve the conductive transmission of energy to the cochlea in otosclerotic ears was only partially successful at closing the air-bone gaps. The mean absorbance at low frequencies in the Surgery group was larger than the other groups in both the ambient test (Figure 4, top) and the tympanometric test at TPP (Figure 5, middle). An increased mean absorbance in the Surgery group would be evidence of either increased transmission through the middle ear to the cochlea or increased internal energy loss within the middle ear. The fact that the air-bone gaps were not closed in the Surgery group relative to the Normal group is evidence that the increased mean absorbance in the Surgery group was likely due to increased energy loss within the middle ear associated with the presence of the middle-ear prosthesis. This finding suggests that absorbance measurements might be useful in improving the effects of corrective surgery for otosclerosis, although further research is needed to explore this idea. A relevant study of fixed stapes pathologies in human temporal bones is that of Merchant et al (2016), which positively assessed the use of WAI to diagnose the particular type of middle-ear pathology related to the presence of conductive hearing loss. A difference between this study using temporal bones with the present study using human participants was that the present study included a group of ears with normal hearing.

An important finding was that the tympanometric reflectance measurement was more accurate than the ambient reflectance measurement in classifying ears as otosclerotic or having normal function. This was based on the univariate classifier results and the construction and evaluation of a multivariate classifier based on the two or three reflectance variables with the largest individual AUCs for classifying ears with otosclerosis. This suggests the clinical potential of using tympanometry absorbance and group delay testing to identify adult ears

with otosclerosis, and possibly in identifying other types of middle-ear disorders. More research is needed with larger numbers of test ears to better understand the strengths and weaknesses of these reflectance tests in diagnosing otosclerosis and other middle-ear disorders.

The ASRTs were compared using the clinical test with its 226-Hz probe tone and the research test in adult ears. In the Normal group, the ASRTs were similar in the clinical and research tests using tonal activators in both ipsilateral and contralateral test modes, and ASRTs were reduced in the research test using BBN activator, which replicates past research (Feeney et al, 2016). Both research and clinical ASRT tests unambiguously classified ears in the Otosclerosis and Surgery groups as having an absent or highly elevated threshold compared with lower ASRTs in the Normal group. As remarked in Keefe et al (2017), the particular promise of the research ASRT test, which is based on measuring shifts in absorbed sound power in the presence of an activator relative to a quiet condition, is the ability to objectively measure ASRTs in infants (Hunter et al, 2017). The operational definition of the ASRT is in terms of the results of the acoustical test used over a finite range of activator levels, and is not intended to represent whether the ASR has a natural threshold. The ability to detect a small ASR shift is limited by system noise and internal noise.

The TEOAE results in the Normal group were similar to normal ears tested in Putterman et al (2017), who used the same TEOAE measurement system as in the present study. The relative absence of TEOAEs in the Otosclerosis and Surgery groups replicated past studies, and extended this feature to higher frequencies (5.7–8 kHz).

Using a test battery approach, a risk for otosclerosis can be estimated on the basis of absent TEOAEs, absent or elevated ASRTs, and an elevated risk of otosclerosis on the multivariate classifier of tympanometric (or ambient) reflectance test variables. A limitation of this finding is that the numbers of participants in the test groups were relatively small, so that future research is needed to evaluate the generalizability of this finding.

CONCLUDING REMARKS

Acoustic WB reflectance testing in the ear canal provides information on middle-ear function that is useful in identifying a risk for otosclerosis relative to normal-hearing ears. A reflectance response is represented using absorbance, which encodes information on reflectance magnitude, and group delay, which encodes information on reflectance phase. Results were compared using a reflectance test at ambient pressure and a reflectance tympanometry test. The tympanometric reflectance was the more accurate test at identifying a risk for otosclerosis. The reflectance in a group of ears after receiving surgery to remedy otosclerosis was different from the reflectance in Normal and Otosclerosis groups of ears. A test battery including reflectance, acoustic reflex, and otoacoustic emission tests provides multiple tests to confirm that a particular adult ear has a high risk of otosclerosis compared with normal-hearing ears.

Acknowledgments. This research was supported by National Institutes of Health grants R01 DC010202 and P30 DC004662. Douglas Keefe has a commercial interest in developing devices to assess middle-ear function and measure TEOAEs. The authors are responsible for the content, which does not necessarily represent official views of the U.S. Department of Veterans Affairs or the National Institutes of Health.

REFERENCES

Allen JB, Jeng PS, Levitt H. (2005) Evaluation of human middle ear function via an acoustic power assessment. *J Rehabil Res Dev* 42(4, Suppl 2):63–78.

American National Standards Institute (ANSI). (2012) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance). ANSI S3.39–1987 (Reaffirmed by ANSI, June 15, 2012).

Cohen J. (1992) A power primer. *Psychol Bull* 112(1):155–159.

Cureoglu S, Baylan MY, Paparella MM. (2010) Cochlear otosclerosis. *Curr Opin Otolaryngol Head Neck Surg* 18(5):357–362.

Feeney MP, Grant IL, Marryott LP. (2003) Wideband energy reflectance measurements in adults with middle-ear disorders. *J Speech Lang Hear Res* 46(4):901–911.

Feeney MP, Hunter LL, Kei J, Lilly DJ, Margolis RH, Nakajima HH, Neely ST, Prieve BA, Rosowski JJ, Sanford CA, Schairer KS, Shahnaz N, Stenfelt S, Voss SE. (2013) Consensus statement: Eriksholm workshop on wideband absorbance measures of the middle ear. *Ear Hear* 34(Suppl 1):78S–79S.

Feeney MP, Keefe DH. (1999) Acoustic reflex detection using wideband acoustic reflectance, admittance, and power measurements. *J Speech Lang Hear Res* 42(5):1029–1041.

Feeney MP, Keefe DH. (2001) Estimating the acoustic reflex threshold from wideband measures of reflectance, admittance, and power. *Ear Hear* 22(4):316–332.

Feeney MP, Keefe DH, Hunter LL, Fitzpatrick DF, Garinis AC, Putterman DB, McMillan GP. (2016) Normative wideband reflectance, equivalent admittance at the tympanic membrane, and acoustic stapedius reflex threshold in adults. *Ear Hear* 38:e142–e160.

Feeney MP, Keefe DH, Marryott LP. (2003) Contralateral acoustic reflex thresholds for tonal activators using wideband energy reflectance and admittance. *J Speech Lang Hear Res* 46(1):128–136.

Feldman AS, Zwislocki J. (1965) Effect of the acoustic reflex on the impedance at the eardrum. *J Speech Hear Res* 8(3):213–222.

Fisch U. (2009) Stapedotomy versus stapedectomy. *Otol Neurotol* 30(8):1166–1167.

Goodman SS, Fitzpatrick DF, Ellison JC, Jesteadt W, Keefe DH. (2009) High-frequency click-evoked otoacoustic emissions and behavioral thresholds in humans. *J Acoust Soc Am* 125(2):1014–1032.

Hannula S, Bloigu R, Majamaa K, Sorri M, Mäki-Torkko E. (2012) Ear diseases and other risk factors for hearing impairment among adults: an epidemiological study. *Int J Audiol* 51(11):833–840.

Herzog M, Shehata-Dieler WE, Dieler R. (2001) Transient evoked and distortion product otoacoustic emissions following successful stapes surgery. *Eur Arch Otorhinolaryngol* 258(2):61–66.

Hosmer DW, Lemeshow S. (2000) *Applied Logistic Regression*. New York, NY: John Wiley & Sons.

Hunter LL, Ries DT, Schlauch RS, Levine SC, Ward WD. (1999) Safety and clinical performance of acoustic reflex tests. *Ear Hear* 20(6):506–514.

Hunter LL, Keefe DH, Feeney MP, Fitzpatrick DF. (2017) Pressurized wideband acoustic stapedial reflex thresholds: Normal development and relationships to auditory function in infants. *J Assoc Res Otolaryngol* 18(1):49–63.

Keefe DH. (1998) Double-evoked otoacoustic emissions: I, measurement theory and nonlinear coherence. *J Acoust Soc Am* 103:3489–3498.

Keefe DH. (2007) Influence of middle-ear function and pathology on otoacoustic emissions. In: Robinette MR, Glattke TJ, eds. *Otoacoustic Emissions: Clinical Applications*. New York, NY: Thieme, 163–196.

Keefe DH, Bulen JC, Arehart KH, Burns EM. (1993) Ear-canal impedance and reflection coefficient in human infants and adults. *J Acoust Soc Am* 94(5):2617–2638.

Keefe DH, Feeney MP, Hunter LL, Fitzpatrick DF. (2016) Comparisons of transient evoked otoacoustic emissions using chirp and click stimuli. *J Acoust Soc Am* 140(3):1949–1973.

Keefe DH, Feeney MP, Hunter LL, Fitzpatrick DF. (2017) Aural acoustic stapedius-muscle reflex threshold procedures to test human infants and adults. *J Assoc Res Otolaryngol* 18(1):65–88.

Keefe DH, Hunter LL, Feeney MP, Fitzpatrick DF. (2015) Procedures for ambient-pressure and tympanometric tests of aural acoustic reflectance and admittance in human infants and adults. *J Acoust Soc Am* 138(6):3625–3653.

Keefe DH, Levi E. (1996) Maturation of the middle and external ears: acoustic power-based responses and reflectance tympanometry. *Ear Hear* 17(5):361–373.

Keefe DH, Ling R. (1998) Double-evoked otoacoustic emissions. II. Intermittent noise rejection, calibration and ear-canal measurements. *J Acoust Soc Am* 103(6):3499–3508.

Kemp DT. (1978) Stimulated acoustic emissions from within the human auditory system. *J Acoust Soc Am* 64(5):1386–1391.

Liu YW, Sanford CA, Ellison JC, Fitzpatrick DF, Gorga MP, Keefe DH. (2008) Wideband absorbance tympanometry using pressure sweeps: system development and results on adults with normal hearing. *J Acoust Soc Am* 124(6):3708–3719.

Margolis RH, Saly GL, Keefe DH. (1999) Wideband reflectance tympanometry in normal adults. *J Acoust Soc Am* 106(1):265–280.

Merchant GR, Merchant SN, Rosowski JJ, Nakajima HH. (2016) Controlled exploration of the effects of conductive hearing loss on wideband acoustic immittance in human cadaveric preparations. *Hear Res* 341:19–30.

Merchant SN, Nadol JB, Jr. (2010) *Schuknecht's Pathology of the Ear*. 3rd ed. Shelton,CT: People's Medical Publishing House-USA.

Metz O. (1946) The acoustical impedance measured on normal and pathological ears. *Acta Otolaryngol* 63(Suppl):3–254.

Mukerji S, Windsor AM, Lee DJ. (2010) Auditory brainstem circuits that mediate the middle ear muscle reflex. *Trends Amplif* 14(3):170–191.

Nakajima HH, Pisano DV, Roosli C, Hamade MA, Merchant GR, Mahfoud L, Halpin CF, Rosowski JJ, Merchant SN. (2012) Comparison of ear-canal reflectance and umbo velocity in patients with conductive hearing loss: a preliminary study. *Ear Hear* 33(1):35–43.

Nikas JB, Low WC. (2011) ROC-supervised principal component analysis in connection with the diagnosis of diseases. *Am J Transl Res* 3(2):180–196.

Plinkert PK, Bootz F, Vossieck T. (1994) Influence of static middle ear pressure on transiently evoked otoacoustic emissions and distortion products. *Eur Arch Otorhinolaryngol* 251(2):95–99.

Putterman DB, Keefe DH, Hunter LL, Garinis AC, Fitzpatrick DF, McMillan GP, Feeney MP. (2017) Assessing sensorineural hearing loss using various transient-evoked otoacoustic emission stimulus conditions. *Ear Hear* 38(4):507–520.

Rappaport JM, Provencal C. (2002) Neurotology for audiologists. In: Katz J, Burkard RF, Medwetsky L, eds. *Handbook of Clinical Audiology*. Philadelphia, PA: Lippincott, Williams and Wilkins, 9–32.

Schairer KS, Ellison JC, Fitzpatrick D, Keefe DH. (2007) Wideband ipsilateral measurements of middle-ear muscle reflex thresholds in children and adults. *J Acoust Soc Am* 121(6):3607–3616.

Shahnaz N, Bork K. (2006) Wideband reflectance norms for Caucasian and Chinese young adults. *Ear Hear* 27(6):774–788.

Shahnaz N, Bork K, Polka L, Longridge N, Bell D, Westerberg BD. (2009) Energy reflectance and tympanometry in normal and otosclerotic ears. *Ear Hear* 30(2):219–233.

Shahnaz N, Longridge N, Bell D. (2009) Wideband energy reflectance patterns in preoperative and post-operative otosclerotic ears. *Int J Audiol* 48(5):240–247.

Shahnaz N, Polka L. (1997) Standard and multifrequency tympanometry in normal and otosclerotic ears. *Ear Hear* 18(4):326–341.

Shanks JE, Shohet J. (2009) Tympanometry in clinical practice. In: Katz J, Medwetsky L, Burkard R, Hood L, eds. *Handbook of Clinical Audiology*. Philadelphia, PA: Lippincott, Williams and Wilkins, 157–188.

Sun XM, Shaver MD. (2009) Effects of negative middle ear pressure on distortion product otoacoustic emissions and application of a compensation procedure in humans. *Ear Hear* 30(2):191–202.

Thys M, Van Camp G. (2009) Genetics of otosclerosis. *Otol Neurotol* 30(8):1021–1032.

Trine MB, Hirsch JE, Margolis RH. (1993) The effect of middle ear pressure on transient evoked otoacoustic emissions. *Ear Hear* 14(6):401–407.

Voss SE, Allen JB. (1994) Measurement of acoustic impedance and reflectance in the human ear canal. *J Acoust Soc Am* 95(1):372–384.

Weatherby LA, Bennett MJ. (1980) The neonatal acoustic reflex. *Scand Audiol* 9(2):103–110.

Werner LA, Levi EC, Keefe DH. (2010) Ear-canal wideband acoustic transfer functions of adults and two- to nine-month-old infants. *Ear Hear* 31(5):587–598.

Zhao F, Wada H, Koike T, Ohyama K, Kawase T, Stephens D. (2002) Middle ear dynamic characteristics in patients with otosclerosis. *Ear Hear* 23(2):150–158.

Zhao F, Wada H, Koike T, Ohyama K, Kawase T, Stephens D. (2003) Transient evoked otoacoustic emissions in patients with middle ear disorders. *Int J Audiol* 42(3):117–131.

Zwislocki JJ. (1962) Analysis of middle ear function. Part I: input impedance. *J Acoust Soc Am* 34:1514–1523.